

ASSESSING THE EFFECT OF INQUIRY-BASED PROFESSIONAL
DEVELOPMENT ON SCIENCE ACHIEVEMENT TESTS SCORES.

Teresa K. Dickson, B.S., M.S.

Dissertation Prepared for the Degree of
DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

December 2002

APPROVED:

James Laney, Coordinator for Curriculum and
Instruction Program and Major Professor

Jane B. Huffman, Minor Professor

Betty Crocker, Committee Member

John Stansell, Chair of the Department of Teacher
Education and Administration

C. Neal Tate, Dean of the Robert B. Toulouse School
of Graduate Studies

Dickson, Teresa K., *Assessing the effect of inquiry-based professional development on science achievement tests scores*. Doctor of Education (Curriculum and Instruction), December 2002, 137 pp., 22 tables, references, 84 titles.

This study analyzed student test scores to determine if teacher participation in an inquiry-based professional development was able to make a statistically significant difference in student achievement levels. Test scores for objectives that assessed the critical thinking skills and problem-solving strategies modeled in a science inquiry institute were studied. Inquiry-based experiences are the cornerstones for meeting the science standards for scientific literacy. State mandated assessment tests measure the levels of student achievement and are reported as meeting minimum expectations or showing mastery for specific learning objectives.

Students test scores from the Texas Assessment of Academic Skills Test (TAAS) for 8th grade science and the biology End Of Course (EOC) exams were analyzed using ANCOVA, chi square, and logistic regression, with the Iowa Test of Basic Skills (ITBS) 7th Grade Science Subtest as covariate. It was hypothesized that the students of Inquiry Institute teachers would have higher scale scores and better rates of mastery on the critical thinking objectives than the students of non-Institute teachers. It was also hypothesized that it would be possible to predict student mastery on the objectives that assessed critical thinking and problem solving based on Institute participation.

This quasi-experimental study did not show a statistically significant difference between the two groups. The effects of inquiry-based professional development may

not be determined by analyzing the results of the standardized tests currently being used in Texas. Inquiry training may make a difference, but because of factors such as the ceiling effect, insufficient time to implement the program, and test items that are intended to but do not address critical thinking skills, the TAAS and EOC tests may not accurately assess effects of the Inquiry Institute. The results of this study did indicate the best predictor of student mastery for the 8th grade science TAAS and Biology EOC may possibly be prior knowledge acquired in elementary school and as demonstrated on the 7th grade ITBS science subtest.

Copyright 2002
by
Teresa K. Dickson

ACKNOWLEDGMENTS

I would like to thank those who have encouraged me in the accomplishment of this study including my husband, Dean, as well as my parents, family and friends. I would also like to acknowledge the technical assistance provided by several talented people: Sue Pritchard with editing, Dae-Yeop Wang and Karen Scott with data analysis, and Mary Arthur, John Doughney, Bryan Reece and Jean Gill with proofreading the final document. Finally, I would also like to acknowledge the motivation provided by the memory of my daughter, Shannon, which challenged me to remain tenacious and focused.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vi
LIST OF FIGURES	viii
Chapter	
1. INTRODUCTION.....	1
Purpose of the Study	
Statement of the Problem	
Hypotheses	
Definition of Terms	
Description of Design	
Significance of the Study	
Organization of the Study	
2. REVIEW OF LITERATURE	16
Introduction	
History of National Standards-Based Science Education	
Development of Texas State Science Standards	
The Standards and Science Instruction	
Inquiry and Science Instruction	
Assessment in Standards-Based Science for Student Achievement	
Equity, Quality and Student Achievement	
Standardized Tests and Accountability for Standards-Based Instruction	
Assessment of Student Achievement in Texas	
Professional Development for Improved Student Achievement	
Staff Development Standards	
Inquiry-Based Staff Development	
Staff Development and Student Achievement	
Conclusion	
3. RESEARCH DESIGN.....	59
Introduction	
Purpose and Statement of the Problem	
Research Hypotheses	

Population and Samples	
Identification of Variables	
Instrument	
Procedures for Experimental Treatments	
Assumptions	
Procedure for Gathering Data	
Conclusion	
Summary	
4. DATA ANALYSIS	78
Introduction	
Collection of Data	
Descriptive Statistics	
Methods of Analysis	
Correlations	
Analysis of Covariance	
Cross Tabulation	
Odds Ratio	
Logistic Regression	
Conclusion	
Summary	
5. DISCUSSION	107
Introduction	
8 th Grade Science TAAS Discussion	
Biology End-Of-Course Exam Discussion	
Prediction of Student Mastery on TAAS and EOC Objectives	
Summary	
Conclusion	
REFERENCES	126

LIST OF TABLES

Table	Page
1. Total Student Population	61
2. Teachers of Student Populations in 8 th Grade Science and 9 th -10 th Grade Biology.....	62
3. Student Populations Divided by Teachers' Participation or Non-Participation in Inquiry Institute	63
4. Intact Class Groups for Quasi-experimental Design.....	64
5. Comparison of Two High School Campuses Showing Equity in Schools	65
6. Sample and State Percentages for TAAS and EOC.....	69
7. Frequency Table for Teacher Attributes	79
8. Frequency Table for Student Attributes	79
9. Correlation Between TAAS and EOC Scale Scores and ITBS Scores	81
10. Mean EOC Scores for Students in Participants' Classes and Students in Non- participants' Classes.....	82
11. Analysis of Variance for Biology EOC Scale Scores Adjusted by ITBS Scores	83
12. Mean TAAS Scores for Students in Participants' Classes and Students in Non-participants' Classes	84
13. Analysis of Variance for TAAS 8 th Grade Science Scale Scores Adjusted by ITBS Scores.....	85

14. Two-way ANCOVA for TAAS Scale Scores using ITBS Covariate for Participation x GT	87
15. Two-way ANCOVA for EOC Scale Scores using ITBS Covariate for Participation x GT	87
16. Odds Ratio for EOC Mastering All Objectives/Teacher Participation in Institute	96
17. Confidence Intervals for Odds Ratios for EOC Participants and Non- Participants Mastering All Objectives	97
18. Logistic Regression for Mastering All EOC Objectives	99
19. Logistic Regression for Pass/Fail EOC Objective 5	100
20. Logistic Regression for Mastering All TAAS Objectives	101
21. Logistic Regression for Pass/Fail TAAS Objective 2	102
22. Summary of Statistically Significant Independent Variables in Logistic Regression Analyses of TAAS and EOC Data	104

LIST OF FIGURES

Figure	Page
1. Student achievement in science depends upon the infrastructure that can support and sustain teacher knowledge and strategies of instruction. (St. John, 2000, p. 3)	73
2. Model for predicting whether a student would pass/fail test objectives and for predicting mastery/ non-mastery of all objectives based on teacher professional development	89
3. Percent of Participants and Non-Participants Mastering All EOC Objectives	90
4. Percent of Participants and Non-Participants Mastering All TAAS Objectives	93
5. Percent of Participants and Non-Participants Pass/Fail EOC Objective 5	94
6. Percent of Participants and Non-Participants Pass/Fail TAAS Objective 2	95

CHAPTER 1

INTRODUCTION

During the past several years, science education reform has introduced factors such as increased academic standards, high stakes testing, and higher teacher accountability. When paired with limited teacher scientific knowledge base, limited instructional resources, and larger class sizes, increased pressure is placed on the states, school districts, schools, and teachers for evidence of ever-rising student achievement in science. These pressures may result in the selection of instructional strategies, which are efficient in dispensing facts and coverage of content, but do little to provide for meaningful investigations. Coverage becomes the focus rather than the development of scientific skills, the construction of a foundation of scientific literacy, or student achievement in science – the intent of the science reform movement (National Research Council [NRC], 1999a). As teachers search for instructional strategies and resources to increase the levels of student achievement and problem-solving abilities, they often resort to the instructional methods and practices used in their own pre-college science instruction. These are based largely on rote memorization of disconnected facts. Many teachers do not have instructional strategies or content background strong enough to give them confidence to teach in an appropriate manner (NRC, 1999a).

The focus of education reform research is the search for programs and practices that make a difference in student achievement. Two seminal studies,

one conducted in 1966 by Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, and York and a second study by Jencks, Smith, Ackland, Bane, Cohen, Grintlis, Heynes and Michelson in 1972 stated that schools make no difference in student achievement and that differences in student achievement are due to factors that schools do not control. These studies differ from more recent findings of the School Effectiveness Research (SER) reported by Reynolds and Teddlie (1999). International research studies attest to the widely held assumption that the “practices that add value” are regularly observable in schools where students achieve and that the transmission of these practices and efforts are part of the policy that improves those schools. According to Fullan (1994), education reform is powerful when teachers and administrators begin working in new ways, discovering that school structures must be altered. The questions of where these reform efforts must begin, top-down or bottom-up, and in what order they should begin are addressed in his research.

In 1996, the National Academy of Sciences (NAS), funded by the National Science Foundation (NSF), and other leading organizations in science education reform published a landmark document entitled the *National Science Education Standards*. The standards described in this document are also described in the American Association for the Advancement of Science (AAAS) document entitled *Benchmarks for Scientific Literacy* (AAAS, 1993). These two documents taken together provide guidance for quality and equity in science education and are the basis for the current science education reform efforts. At the heart of the *Standards* and *Benchmarks* are inquiry-based learning and hands-on

investigations grounded in the theories of Vygotsky, Piaget, Bruner, and other constructivist and cognitive-learning theorists. Research by the Inverness Research Associates concludes that when inquiry is the focus of instruction, students are able to expand their knowledge, recognize what they don't know, and become fearless in going beyond the boundary of what they do not know in gaining scientific knowledge and skills. Student achievement is positively impacted through inquiry-based instruction (St. John, 2000).

The National Staff Development Council (NSDC) guidelines suggest that quality professional development address the context, content, and processes that are needed to close the gap between national and state instructional standards and student achievement when evaluated for mastery of those standards. In order for science teachers to become proficient in the best practices of quality and equitable science instruction for the success of all learners, professional development must be presented in the context of best practices for adult learners, using the process of scientific inquiry, and based on the content of the national and state benchmarks for scientific literacy. Based on the guidelines established by the *National Science Education Standards* (NRC, 1996), professional development for teachers requires learning essential science content through the perspectives and methods of inquiry.

As teachers experience inquiry in staff development models designed as inquiry institutes, they learn and experience how best to use different kinds of hands-on instruction. "It also gives them insights into the learning processes as they experience the same feelings of optimism, frustration, competitiveness, and

potential for learning their students feel in each of these situations” (Kluger-Bell, 2000, p. 40).

Purpose of the Study

It is the purpose of this study to examine the academic achievement of students whose teachers participated in an inquiry-based professional development institute and the academic achievement of students whose teachers did not participate in an inquiry-based professional development institute to determine whether a relationship exists. Simply stated, the problem of this study is to determine the relationship between the participation of 8th grade science and biology teachers in an inquiry-based professional development institute and student achievement as measured on the state mandated achievement tests, the 8th grade science Texas Assessment of Academic Skills (TAAS) and the biology End of Course (EOC) exam for the 9th and 10th grade students enrolled in biology.

Statement of the Problems

The problem of this study is to determine to what extent a relationship exist between the participation and non-participation of science teachers in an inquiry-based professional development institute (guided by instructional and professional standards) and student achievement as measured on the state mandated (standard-based) achievement tests.

It is also to determine if it is possible to predict student success or failure in passing specific objectives that assess critical thinking skills, and for mastery of all objectives on state mandated achievement tests, based on teacher

participation or non-participation in an inquiry-based professional development program entitled the Inquiry Institute.

Hypotheses

It was hypothesized that the 8th grade science TAAS scores for students of Inquiry Institute teachers had:

H₁: higher mean scale scores than students of non-Inquiry Institute teachers.

H₂: higher passing rates on objective 2 than students of non-Inquiry Institute teachers.

It was also hypothesized that 9th and 10th grade biology EOC exam scores for students of Inquiry Institute teachers had:

H₃: higher mean scale scores than students of non-Inquiry Institute teachers.

H₄: higher passing rates on objective 5 than students of non-Inquiry Institute teachers.

Finally, it was hypothesized that students in identified subgroups, when taught by Inquiry Institute teachers had:

H₅: higher TAAS and EOC objective mastery rates than when taught by non-Inquiry Institute teachers.

Definition of Terms

Benchmarks for Scientific Literacy - Included in Project 2061's second publication of the same name (1996), these standards are descriptions of how students should progress towards science literacy, recommending what they

should know and be able to do by the time they reach certain grade levels.

Together, with the *National Science Education Standards*, the two publications are intended to help guide reform in science, mathematics, and technology education.

Biology End of Course Exam (EOC) – This standards-based assessment measures mastery of the statewide curriculum for high school biology courses in order to ensure that high academic standards are being met. Demonstrating satisfactory performance on EOCs became an additional means for students to be eligible to graduate beginning in the 1998-1999 school year.

Brain-Based Learning –This term has been coined by educators who advocate changes in teaching methods based on neuroscientific research about brain function and dendrite growth. Neurologists, as a result of functional Magnetic Resonance Imaging (MRI) and Computerized Axial Tomography (CAT Scans), are now able to watch patterns of activity within the cerebral cortex, causing scientists to revise earlier assumptions about how individual learning actually occurs (Williams, 1999). Some of these practices include enrichment of the learning environment, connecting new learning to prior learning, favoring hands-on and participatory learning, as well as using novelty to increase memory when presenting new material. These brain-based strategies have the potential to increase the retrieval of the information from long-term memory (Jensen, 1996).

Cultural-Historical Theory of Learning - Based on the research of Lev Vygotsky, this theory states that learning is socially mediated and meaning is

constructed as it is gained through the interactions with the tools, symbols, words and people in a culture. As children use tools in their investigations into the nature of science, their insights are made meaningful when the teacher provides the context for the knowledge gained and helps the child tie the new learning to prior knowledge (Gredler, 1976). Vygotsky (1978, cited in NRC, 1999c) described this assistance as the Zone of Proximal Learning.

Discovery Learning – Based on the research of theorist Jerome Bruner, insight in education is more important than memorization. Caine and Caine studied Bruner's theory describing meaning based on insight and marked with a sense of relief and energy, such as when a picture that is made up of a collection of dots "suddenly" turns into a recognizable picture (Bruner, Goodnow & Austin, 1967, cited in Caine & Caine, 1994, p. 103). Bruner's theory held that students construct knowledge through experiences that continually add to their body of knowledge and learning. As advocated by Bruner, there is a need to allow students to restructure their thoughts and beliefs about a concept and to incorporate new ideas into connections with what they already know. Sometimes they must discard long-held beliefs that were based on misinformation or distorted views. According to Joyce and Weil (2000), this approach is also called constructivism.

Equity and Quality in Science Instruction – This set of recommendations, based on *Science for All Americans* (AAAS, 1989), constitutes a common core of learning in science, mathematics, and technology for all young people, regardless of their social circumstances and career aspirations. In particular, the

recommendations pertain to those who in the past have largely been bypassed in science and mathematics education: ethnic and language minorities and girls (AAAS, 1989).

Hands-On Learning - Students are given the opportunity to interact with the materials of investigation rather than receive all the instruction through lecture or reading. Hands-on learning can take many forms. According to Rankin, "All hands-on is not inquiry and not all inquiry is hands-on. Using hands-on methods does not always ensure effective science teaching, nor is it necessarily indicative of an inquiry-based approach" (2000, p. 34). Rankin further states that hands-on learning is distinguished by the "amount of flexibility a teacher allows in order for children to develop individual curiosity and ways to solve problems. This is different from a situation in which a teacher poses a question and then directs all the students to take the same pathway to find a common solution" (p. 35).

Inquiry-Based Science – This approach to instruction includes two methods, general inquiry and scientific inquiry. In general inquiry instruction, science process is the focus, content moves to the background. The students conduct experiments and construct their own meaning of the events and phenomena that occur naturally. It is through investigations at the students' own rates and levels of ability that learning takes place (Chiappeta, 1997). In scientific inquiry, content becomes the focus, and process moves to the background. Through the use of questioning strategies, as students investigate and conduct experiments, the teacher is able to direct and channel the learning into an understanding of the larger concepts and principles of science that

explain the phenomena. Students then apply these newfound skills and insights to new situations with an understanding of how the larger concept relates and connects to the smaller ones (Kluger-Bell, 2000).

Intellectual Development Theory – As described by Joyce and Weil (2000), Piaget studied how children mature intellectually and begin to make meaning of their experiences. As young children begin to think and organize their thought processes into concepts and intellectual structures, experiences are assimilated into their structures of thought and patterns of behavior called schemas. With more experiences mediated by the schema, the child is able to accommodate the experiences, give them meaning, and gain an understanding of how it all fits together. As the experiences become more complex, the schema become more complex, and the child's intellectual capacity grows. The child understands more and moves from concrete to abstract thought.

National Assessment of Education Progress (NAEP) - According to Schrag, (1997) the NAEP is better known as the Nation's Report Card; it is widely considered to be among the most reliable measures of academic achievement to show the impact of changes in demographics, ethnic populations, and socioeconomic factors on student achievement. The test includes hands-on tasks that probed students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills. There were multiple-choice questions and constructed-response questions. The areas of science covered by the NAEP were earth science, life science, and physical science. The results are intended for teachers and relate directly to

students' performance, classroom practices and school climate (NAEP, 2000).

National Science Education Standards– Included in a publication by the same name, they offer a coherent vision of what it means to be scientifically literate, describing what all students, regardless of background or circumstance, should understand and be able to do at different grade levels in various science categories. The standards address the practice of science teaching to provide students with experiences that enable them to achieve scientific literacy. These standards reflect the principles that learning science is an inquiry-based process, that science in schools should reflect the intellectual traditions of contemporary science, and that all Americans have a role in improving science (NRC, 1995).

Professional Development – Also known as staff development, these efforts are judged by their contribution to student learning. It is based on the premise that educators must continually increase their knowledge and improve their skills in order to assure higher levels of achievement and student learning. Guskey (2000) states that as a process it is intentional, ongoing, and systemic. It is intended to bring about positive change and improvement, guided by a clear vision of purposes and planned goals. The goals are the criteria used to select the content, materials, procedures, and assessments used.

Professional Development Standards – As established by the National Staff Development Council (2001), the standards are organized into three categories: context, process, and content. These ensure that staff development is designed to make an impact on student learning and achievement. According to the NRC (1996) "Professional development must have in-depth, ongoing,

learning opportunities to learn science through inquiry; integrate knowledge of science, learning, and teaching; engage in continuous reflection and improvement; and build coherent, coordinated programs for professional learning” (p. 12).

Data-Driven Instruction –Disaggregated assessment data gathered either from standardized tests or other evaluation methods reveal the gap in areas where students need additional assistance in order to advance to their maximum academic potentials. Data disaggregation is the act of taking test items and breaking them into smaller components, skills, knowledge, and content for teaching in smaller pieces. This allows for adjustments to the curriculum or the work plan so that changes in teaching are the result. According to English (2000), “such changes may include or exclude different content, may spend more time on certain areas to teach, and may alter the scope and/or sequence of curricular content” (p. 19).

Schools Effectiveness Research (SER) – According to the collection of education research focused on school reform efforts, edited by Teddlie and Reynolds (1999), there are three major strands of school effectiveness research. In the first strand, the outcomes are limited to student achievement on standardized tests. It is based on research studies by Coleman in 1966 and Jencks in 1972 that link student achievement to family socioeconomic status and other factors outside the school setting (Marzano, Pickering & Pollock, 2001). The second strand, based on a collection of education research, supports the argument that *how* content is taught is equally as important as *what* is taught.

The effective schools literature reveals strategies that empower schools and districts to implement the long term, systemic changes needed to ensure that every student succeeds (Lezotte, 2002). The third strand is based on research that summarizes strategies for guiding the systemic development of reform policy (Fullan, 1994, Teddlie & Reynolds, 1999).

Science Literacy – Literacy encompasses mathematics and technology as well as the natural and social sciences. Among its many facets it includes being familiar with the natural world and respecting its unity; understanding key concepts and principles of science; having a capacity for scientific ways of thinking; and being able to use scientific knowledge and ways of thinking for personal and social purposes, (AAAS, 1989).

SISS Study – Second International Science Study was a comparative study conducted in China and the United States in the mid 1980s by the International Association for the Evaluation of Education Achievement (IEA). According to Wang (1996) empirical data was used to evaluate the science competence of United States students against the international standards of industrialized countries and became one of the catalysts for the science standards reform movement.

Texas Assessment of Academic Skills (TAAS) – The TAAS is the state mandated assessment for mastery of the state standards for students in grades 3-12. Eighth grade students are TAAS tested in writing in February and math, reading, science and social studies in April.

Texas Essential Knowledge and Skills (TEKS) – The TEKS are grade-

level and subject-specific standards and student expectations that comprise the curriculum for Texas public school students. The TEKS are written for grades K-12 and specify skills, knowledge, and processes that students are to know and be able to do at each grade level. These standards for the education of Texas' students were passed by the Texas Legislature in 1997, were to be implemented by September 1998, and are assessed by the TAAS.

TIMSS Study - The Third International Math and Science Study conducted during the 1994-95 school year has been used extensively to compare the mathematics and science achievement of students and the instructional practices of schools worldwide. According to the Eric Digest (2000), students in public and private schools were tested at three grade levels: fourth (nine years old), eighth (thirteen years old), and twelfth (final year of secondary schooling). Tests reflected educational goals and standards very similar to our national standards and assessed student achievement in 23 countries. The test was repeated in 1999, called the TIMSS – R, focusing on the mathematics and science achievement of eighth graders with 38 nations participating. The advantage of the TIMSS-R is that it allowed the United States to compare the achievement of its original TIMSS fourth grade cohort as 8th graders to its original 8th graders four years later (EDO-SE-00-05, 2/24/02).

Description of Design

This study investigates whether there are differences in students' achievement test scores when their teachers participate in a professional development program using the Immersion into Science model described by

Loucks-Horsley, Hewson, Love, and Stiles (1998). During the 2001-2002 school year, 42 teachers, who teach science in grades kindergarten through twelve, participated in a four-day Inquiry Institute, which is part of the school district's Science Initiative. For the purposes of this study, quantitative research methods were utilized. Student achievement data are analyzed from the 8th grade science TAAS given on April 19, 2002 and the biology EOC exam given during a seven-day window starting May 13, 2002. The Iowa Test of Basic Skills (ITBS) from the seventh grade science subtest was used as a covariate for prior learning. This study attempted to establish a cause/effect relationship between student achievement and two models of teacher staff development; one of which models inquiry-based instructional strategies and the second which models the traditional staff development format.

Significance of the Study

The findings of this study are significant because they add to the body of knowledge concerning the use of inquiry as a process for teacher professional development and the effect of professional development on student achievement as measured on a standards-based assessment.

Organization of the Study

Even though it is stated in the *National Science Education Standards* (NRC, 1996) and the TEKS (TEA, 1998) that students are to be engaged in inquiry-based investigations, teachers appear to be unconvinced of the effectiveness and efficiency of inquiry-based instructional strategies especially when the pressures for accountability in student achievement come into play.

This chapter has presented an introduction to the inquiry-based standards at the state and national levels, the purpose of the study, a statement of the problem, and the hypotheses that it will statistically analyze for significance. The terms used in the study were defined, a description of the study was given, and the organization of the study was presented. Chapter 2 will provide a review of the literature.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

In this chapter, the review of literature begins with a brief history of the national standards-based science education and the development of state standards for science. Literature concerning instructional standards and the use of inquiry, as an instructional strategy and process, are reviewed. To provide background to study how equity and quality affect student achievement and accountability, research on the international, national, and state-level assessment programs are reviewed. Finally, in this chapter, is a review of research concerning science professional development and the use of inquiry in the process, context, and content of science professional development programs as found in the literature. It is the intent of this chapter to develop a rationale for further study of the effects of inquiry-based professional development on student achievement as measured by state mandated standardized tests.

The History of National Standards-Based Science Education

Student achievement in science has become a major focus of education in the United States. With the launch of the Soviet Union's satellite, Sputnik I, in the 1950's and the concern for equity brought about by the civil rights movement of the 1960's, the public education system in the United States became the topic of parental, local, state, and federal scrutiny. In the 1970's, dropping SAT and other performance standard scores helped to galvanize the call for "back-to-basics"

education. Against this backdrop, a mandate for learning expectations beyond the minimum competencies was issued. In 1983, the federal government provided the stressors on the American educational system that are still affecting policy decisions 19 years later. The National Commission on Excellence in Education (NCEE) released a report in 1983 entitled *A Nation at Risk* that warned the “educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people” (1983, p. 5). According to Goldberg and Harvey, two staff members of the NCEE, the report “created more furor in United States education than any event since Sputnik I. America is at risk, they say, but there is a cause for optimism, too” (1983, p. 14).

Since the release of *A Nation at Risk* in 1983, several major events impacting science education have occurred. Project 2061 was launched in 1985 by the American Association for the Advancement of Science (AAAS) to build capacity for the improvement of science, math, and technology education. In 1986, the SISS Study was conducted to measure and compare student achievement with students from other nations. Based on preliminary data, students from the United States did not rank as highly, in spite of some initial reform efforts. When the official results for the SISS were released in 1992, the ranks for the United States were low but the students’ scores were close to the national average on other assessments. Reaction in the national media rekindled and reinforced the national mandate for standards-based reform (Bracey, 1997).

Ravitch, (Bracey, 1997) as a member of the original 1983 NCEE, stated

that standards would improve achievement by clearly defining what is to be taught and what kind of performance is expected. In studies conducted on data from the 1980's norm-referenced SAT tests, the results indicated a declining level of knowledge in 11th grade students in both American history and literature. Ravitch (2001) stated that the causes for the decline in student achievement, which led to the standards-based reform movement, were:

First, the belief that schools should be expected to solve all of society's problems; second, the belief that only a portion of children need access to a high-quality academic education; and third, the belief that schools should emphasize students' immediate experiences and minimize (or even ignore) the transmission of knowledge. (p. 465)

Ravitch maintained "schools cannot succeed unless they focus on what they do best...they cannot be successful as schools unless nearly all of their pupils gain literacy, numeracy, as well as a good understanding of history, and the sciences, literature, and a foreign language" (p. 465). She goes on to state, "To be effective, schools must concentrate on their fundamental mission of teaching and learning. And they must do it for all children. That must be the overarching goal of schools in the twenty-first century" (p. 467). Conclusions drawn from her study and a similar one by Whittingham in 1992 (Bracey, 1997) were used to stir public sentiment and became part of the momentum for reform efforts in public education.

In the *Sixth Bracey Report on the Condition of Public Education*, Bracey (1996) refuted the research of Ravitch and Flinn (1987) and the general outcry following the publication of *A Nation at Risk* as a false crisis. His report was based on the data from the study released by the National Center for Education

Statistics in 1992, showing the TIMSS scores remained about the same or slightly improved over the previous SISS study.

In 1989, the President and National Governors' Association adopted national educational goals at the First National Education Summit. These goals were part of the Goals 2000: Educate America Act, Title III, Sec. 302, which became law in 1994. It provided funds as policy levers for states to improve and standardize their school systems with programs to meet their specific needs (Department of Education, 1998). There are eight National Education Goals. According to Wertz (1999), each goal was based on commissioned papers written by members of the National Education Goals Panel. Lauren Resnick, as one of the original members of the National Education Goals Panel wrote goal number 3:

#3 Student Achievement and Citizenship – By the year 2000, all student will leave grades four, eight, and twelve having demonstrated competency over challenging subject matter including English, mathematics, **science**, [bold added for emphasis] foreign languages, civics and government, economics, art, history, and geography, and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our Nation's modern economy (pp. 14-15).

Commissioned by the National Education Goals Panel for her expertise in science and mathematics excellence standards, Senta Raizen wrote goal number 5:

#5 Mathematics and Science – By the year 2000, the United States will be first in the world in mathematics and science achievement.

- Mathematics and science education, including the metric

system of measurement, will be strengthened throughout the system, especially in the early grades.

- The number of teachers with a substantive background in mathematics and science, including the metric system of measurement, will increase by 50 percent.
- The number of United States undergraduates and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering, will increase significantly. (pp. 18-19)

It was the expectations of the Goals 2000 Act that there would be a layering of the standards at the national, state, and local levels. The national standards would define broad skills and concepts in curricular fields, and be voluntary. The NSF and AAAS provided the guidance for the national science standards. The development of standards is described in the Project 2061's book, *Science for All Americans* (AAAS, 1989), which consists of a set of recommendations on what understandings and ways of thinking are essential for all citizens in a world shaped by science and technology.

The *Benchmarks for Science Literacy*, published in 1993 by the AAAS, (from here on referred to as *Benchmarks*) is Project 2061's second publication, and specifies how students should progress towards science literacy as they reach certain grade levels. *National Science Education Standards* (referred to as *Standards* from this point forward) were written through the efforts of many national organizations and published in 1995 by the National Research Council (NRC). Primarily the NSF funded the document. It offered a coherent vision of what it means to be scientifically literate, describing what all students, regardless of background or circumstance, should understand and be able to do at different

grade levels in general science categories. Both the *Standards* and the *Benchmarks* addressed the exemplary practice of science teaching that provides students with experiences that enable them to achieve scientific literacy.

According to Ravitch (2000), because of the commitment of thousands of people, there is a realization of what students are capable of doing. Glickman (1998) summarized the “restructuring” period of education, from the late 1980s to the late 1990s, as a time that generated the largest and most sustained rethinking of schools along the lines of democracy, with the inclusion of all students as “active, curious, and wise citizens” (p. 46). Looking now from those beginnings, the question is posed: “How do we sustain such work when recent history indicates that it will be curtailed?” (p. 46) As stated in the Epilogue of the *Standards*

With distributed leadership and coordinated changes in practice among all who have responsibility in the reform of science education reform, advances in science education can rapidly accumulate and produce recognizable improvement in the scientific literacy of all students and citizens. Recognizing the challenges these standards present, we encourage Legislators and public officials to strive for policies and funding priorities aligned with the National Science Education Standards. (pp. 244-245)

The Development of Texas State Science Standards

State standards are specific content statements of what students should know and be able to do. The national standards are used as a basis for the development of statewide standards in 47 of the 50 states plus the District of Columbia and Puerto Rico (Department of Education, 1998). The standards then become the frame for assessment of student mastery. State policy makers decide the assessment standards but local leaders decide curriculum standards,

the sequence of study and rationale. Local educators tailor the national and state standards, making them more specific in instructional areas and describing what should take place in the classroom. Assessment of student achievement towards mastery of these standards is designed and monitored at the state level for standardization and accountability.

According to Vornberg (1998), the Essential Elements, the first standards for education in Texas (Chapter 75 of the Texas Education Code), were developed in response to legislation passed in 1981 and adopted by the State Board of Education in 1984. The Texas Essential Knowledge and Skills (19 Texas Administrative Code, Chapter 74, 1998) are part of the rules governing Texas' schools. In 1997, the 74th Texas Legislature passed Senate Bill 1, which called for the State Board of Education to adopt the Texas Essential Knowledge and Skills, known as the TEKS, for the state required curriculum.

The Science TEKS for grades K-12 are the standards and objectives for the science curriculum. According to the Texas Statewide Systemic Initiative Issue Brief (Dana Center, 1998), the “TEKS provide an opportunity for looking at curricula differently” (p. 1). School districts are examining current curriculum documents and are working to bring the curriculum in line with the state standards. Based on the results of the international studies, a review of trends in current national curriculum reform efforts, expectations of the community, and assistance by state and professional organizations, curriculum must reflect these views. Districts and schools are expected to ensure implementation of the TEKS, which changes the nature of local curriculum development. “Success will be

gauged by student achievement on TEKS-based measures such as the revised TAAS and End-of-Course tests.” (p. 2)

The SSI has identified several key actions that districts and schools should take to successfully implement the TEKS. These recommendations are based on their knowledge of best practice and research regarding the relationship between effective curriculum development and related student achievement (1998).

1. Providing opportunities and a structure for all teachers to understand the TEKS by studying them in depth both at their grade level and across grades, with a focus on what is unique and new about the TEKS for any particular topic at any particular level;
2. Making important decisions, based on this study of the TEKS, about the philosophy and direction of mathematics and science instruction, including the role of technology, types of instructional groupings, the use of hands-on materials, etc.;
3. Providing teachers within a grade or for a course the opportunity to clarify the focus, group TEKS statements for instructional effectiveness, and define an instructional sequence for the grade or course...as a school or district level activity...
4. Analyzing and selecting curriculum programs that support the TEKS ...and developing implementation plans for the curriculum;
5. Providing opportunities for meaningful, ongoing professional development for teachers, administrators, and other appropriate individuals;
6. Providing necessary calculators, hands-on materials and instructional resources to students and teachers;
7. Developing and implementing policies that support sound mathematics and science instruction... (1998, p. 2)

The TEKS reveal outcomes of K-12 science education as encompassing scientific processes in students’ abilities to conduct field and laboratory investigations that are safe, environmentally appropriate, and ethical; to use scientific methods during field and laboratory investigations; and to use critical

thinking and problem solving to make informed decisions (19 Texas Administrative Code (TAC), Chapter 74, 1998).

The TEKS are divided into specific content strands that spiral and build from kindergarten through 12th grade, identifying the processes, skills, and content knowledge for each. The science strands include living systems, earth systems, space systems, matter and energy, force and motion, scientific processes, and data analysis. The contexts for the strands are the use of inquiry; properties, patterns, and models; constancy and change; and systems.

The Texas Education Agency adopted the TEKS in September 1997, and with the final revision, based on the feedback from many sources, they were published in 1998. Implementation began in all Texas public school classrooms in the fall of 1998. Many district-constructed curriculum documents contain differentiation for the gifted and talented students and instructional consideration for students with special needs. As the curriculum revisions occur, teachers are reminded of the use of sound instructional practices and alignment to the revised test, which is to be released in 2003. The revised curricula are to include technology based labs, skill-based assessments as well as inquiry processes to meet the standards, and to build towards student mastery of the TEKS-based assessments (Dana Center, 1998).

The Standards and Science Instruction

The conclusions of learning theorists such as Vygotsky, Piaget, Bruner, and advocates of brain-based learning such as R.N. Caine, G. Caine, and Jensen, are key to the instructional processes and strategies found in the

Standards, Benchmarks, and the TEKS. The framework established by the standards moves the learning of children in grades K-12 from concrete to abstract, as they gain experience and mature. Concrete, hands-on experiences with models, materials, and equipment are effective in providing ways for students to make meaning and explain abstract scientific phenomena, which can be outside a student's ability to reason or understand.

Research reviewed by Wadsworth (1978, reported in Joyce, 2000) included studies by Klausmeir and Hooper (1974) and Kohlberg (1977). Wadsworth provides rationale for curriculum matched to student levels of development and instruction that accelerates intellectual development. Inquiry-based instruction uses such a framework, which adjusts instruction to learner developmental stages. Wadsworth explained that teaching is the creation of environments, which give students practice with particular operations. Student roles must be active and self-discovering. Teachers' roles include organizers of the environment, assessors of children's thinking, and initiators of group activities. Research to determine the effectiveness of the teaching models, over the past 35 years as compiled by Joyce (2000) indicated generally positive results.

Inquiry and Science Instruction

In an overview of the *Standards*, inquiry is identified as an instructional and educational standard. "Learning science is something that students do, not something done to them. 'Hands-on activities', while essential, are not enough. Student must have 'minds-on' experiences as

well” (NRC, 1996, p. 2). As stated in the 9th through 12th grade inquiry standard of the *Standards*

For students to develop the abilities that characterize science as inquiry, they must actively participate in scientific investigations, and they must actually use the cognitive and manipulative skills associated with the formulation of scientific explanations. (p. 173)

Science programs exist in all schools; however, the qualities of these programs vary. The key components to programs that promote student inquiry as identified by Edwards (1997) include keeping the focus on real-life experiences of the students. He points out that when examining current science curricular materials whose primary aim is inquiry, a basic flaw is revealed:

It is difficult to create materials that in fact provide inquiry experiences for each student...it is hard to create materials that attend to the personal real-life experiences of students... To have bona fide inquiry experiences, students must formulate their own questions, create hypotheses, and design investigations that test their hypotheses and answer the questions proposed. Published materials are generally too structured to provide the necessary freedom for students to engage in these important inquiry skills. (p. 19)

According to the SSI (Dana Center, 2002b), when identifying the elements of an effective science program, the TEKS clearly state that students should be

Involved in inquiry-based instruction. Their experiences include hands-on, minds-on learning. Students actively learn how to observe, ask questions, plan investigations, gather information using tools, predicts, propose explanations, communicate results, and reflect on the processes they have used. Students learn how to think critically, obtain information, think for themselves, and actually learn how to learn. (pp. 1-2)

According to Schmidt (1999), teachers studied the use of inquiry in developing classroom literacy learning strategies. Using a modified version of the K-W-L format published by Brian in 1998, teachers in grades K-5 followed the

procedure not only for building upon prior knowledge but also as a means for recording and generating questions throughout the study. To develop an understanding of inquiry, they first attended a staff development workshop where they learned about the constructivist approach to learning and inquiry strategies. They learned that with inquiry if children are “presented with first hand experiences to experiment, solve problems and discover how the world functions...they question, plan, investigate, reflect, explain, and summarize.” (p. 789)

After the workshop they began developing inquiry lessons. Realizing that students’ questions actually define what they do not know about the topic, the teachers developed a framework of their own for question generation, “K-W-L-Q; What I know, what I want to know, what I learned, and more questions” (p. 789). Using this framework, students perform inquiries and generate hypotheses that can be tested. In this manner, students locate new information and data, answer their own questions about the natural world, and solve problems with an appropriate level of assistance from the teacher.

Inquiry science requires hands-on experiences. The National Science Teachers Association (1990), in a position statement about laboratory science, offers the following standards for time to be spent in laboratory and field investigations:

- Preschool/Elementary science classes must include activities-based, hands-on experiences for all children. Activities should be a minimum of 60% of the science instructional time.
- All middle level and high school science courses must offer laboratory experiences. There should be a minimum of 80% of the

science instruction for middle school.

- 40% at the high school level should be spent in laboratory-related experiences; this time includes pre-lab instructions in concepts relevant to the lab, hands-on activities by the student, and a post-lab period involving communicating and analysis. Computer simulations and teacher demonstrations are valuable but should not be substitutions for laboratory activities. (NSTA, 2001)

The TEKS for high school science require that laboratory investigations and field experiences make up 40% of instructional time. Although there is no state mandated requirement for laboratory time for the middle school level, the TEKS require extensive use of inquiry and hands-on investigations for mastery of the science knowledge and skills (1998).

In a seminal four-year study from 1995 – 1999, Klentschy, Garrison, and Amaral (1999) compared the use of kit-based science instruction and student achievement in schools in the Valle Imperial Project in California. “The program had five critical elements: high quality curriculum, sustained professional development and support for teachers and school administrators, materials support, community and top level administrative support, and assessment” (p. 3).

The students in selected pilot schools used four kits per year in grades 1-5 and three kits in kindergarten. Students were engaged in process skill development and in-depth knowledge building through the hands-on opportunities provided by the kits-based instruction. The Science Section of the Stanford Achievement Test, 9th Edition, Form T was used to assess all 4th and 6th grade students in the entire district. The 9th edition is constructed to mirror the philosophy of *Science for All Americans* (NCR, 1989). “The results of the study were favorable towards the kit-based instruction...reporting a 14-percentile point

difference. Also found noteworthy were improvement for females, economically disadvantaged, and minority students” (p. 6).

In a review of instructional programs by ERIC Clearing House, Landis concluded that hands-on activities and inquiry foster meaningful learning in the science classroom (ERIC, 1996). These same activities overcome gender bias. Solving problems, which are rooted in real-world observations has been described as an influential factor in the decision-making of girls who elect to study science (Harding, 1985, cited in ERIC Digest, 1996, p. 2). Also, specific steps can be taken to remove gender bias and to include the thinking of scientists with diverse backgrounds, overtly directed toward retaining the interest of women in pursuing science related careers (Rosser, 1993, as cited in ERIC Digest, 1996, p. 2).

Assessment in Standards-Based Science for Student Achievement

“In the vision of the *Standards*, assessment data provides students with feedback on how well they are meeting the expectations of their teachers and parents, teachers with feedback on how well their students are learning, districts with feedback on the effectiveness of their teachers and programs, and policy makers with feedback on how well policies are working” (p. 76). Also, as districts compare their programs with national standards and their local vision, an increased number of districts will need to take into account statewide standards of learning that are reinforced by mandatory statewide requirements. It is suggested that communities may decide to upgrade their programs dramatically to reflect innovations that will allow for student success on these assessment

instruments (NRC, 1999a).

Schools are feeling a tremendous push to improve student achievement levels and the quality of the teaching force (Brendt, 2000) based on the new standards-based programs advocated by *A Nation At Risk*. Because of the need for greater accountability for student learning and for generating empirical proof that American schools are meeting the desired outcomes, local policy makers have been encouraged by national policy organizations to do the following to insure higher levels of student achievement on standardized testing measures:

1. add to the curriculum in the areas of science, math and technology;
 2. lengthen the time students spend in school;
 3. increase the requirements for school graduation; and
 4. develop and implement more rigorous assessments of student learning.
- (p. 5)

Popham writes “when policymakers create accountability systems centered on students’ test scores, they assume higher scores reflect better instruction” (2001, p. 1). Unfortunately, the pressure on teachers to prepare students for high-stakes multiple-choice achievement tests greatly impacts instruction. The need for the efficient dispensing of facts and coverage of content forces the use of instructional strategies that do not provide meaningful investigations, the development of scientific process skills, the construction of a foundation of scientific literacy, or true student achievement in science (NRC, 1999b).

Popham poses that the “whole strategy of basing instructional evaluations on shifts in students’ test scores depends on the proposition that differing levels of instructional effectiveness will produce related levels of student test

performance” might not be defensible. He also questions if making year-to-year comparisons of test scores, based on the fluctuation of student caliber from year to year, is defensible (2000, p. 1). Popham reasons that the standardized tests often are not matched to assess the content being taught and as such are not valid assessments of instruction. Secondly, he states that achievement tests measure a variety of things - some of which the student did not necessarily learn at school:

1. what the student learned in school;
2. the students’ socioeconomic status;
3. the students’ inherited academic aptitude. (p. 2)

Stated in the overview of the Assessment Standard in the *Standards*, “feedback does lead to changes in the science education system by stimulating changes in policy, guiding teacher professional development, and encouraging students to improve their understanding of science” (p. 76).

Equity, Quality and Student Achievement

As Brandt (1998) discussed the changing context of teaching, he pointed out that society is being transformed as we move toward an information society and global economy. People outside the school community are pressing people inside to teach students expanded processes, skills, and content. Businesses, industry, and institutions of higher learning are pressing people inside the school system to develop a multicultural awareness, to teach basic skills, as well as to teach how to frame and solve problems. These demands have strong implications for teaching and add stressors to an already challenged system, and are expected much faster than the system can accommodate.

Emerging demographics also stress the system. The number of poor families has risen dramatically. Although, according to Glickman (1998), students with social, physical, or educational problems have always been a challenge to their teachers, reform efforts to increase their level of achievement on measures of standard mastery to inappropriately high levels may prove more than the system can handle. Gaps in student achievement in these under-served populations, compared to the better-served populations, are widening, creating what many perceive as a threat to democracy, and placing our nation at risk.

A research study concerning the effects of instruction on student achievement for the Equality of Educational Opportunity Report was published in 1966 with Coleman as senior writer. Marzano, Pickering and Pollock of McREL (2001) reviewed this study as well as a similar study by Jencks in 1972, as part of their research on student achievement. The McREL study found these reports to be of seminal importance not because of their findings, but because of the “flaws in their conclusions” (p. 2).

Coleman analyzed data from 600,000 students and 60,000 teachers in more than 4000 schools. Using regression analysis that mixed levels of data analysis, he was able to conclude, “the quality of schooling a student receives accounts for only about ten percent of the variance in student achievement” (p. 1) and “the majority of the differences in student achievement can be attributed to factors like the student’s natural ability or aptitude, socioeconomic status of the student, or the student’s home environment” (p. 2). The research by Jencks et al. (1972) corroborated those of Coleman et al. “Most differences in ... test scores

are due to factors that schools do not control” (p. 2). In a review of Coleman’s study, Teddlie and Reynolds (1999) noted that some of the factors studied by Coleman related to school resources which were not related to school achievement, such as per pupil expenditures, school facilities, and numbers of books in the libraries. They stated the findings are still widely accepted (Teddlie and Reynolds, p. 10). The alternate conclusions to the studies of Coleman and Jencks are further discussed in the Literature Review section concerning standardized assessments.

In the *Sixth Bracey Report on the Condition of Public Education* (1996), Bracey refutes the research of Ravitch and Flinn (1987) and the findings for *A Nation at Risk*. Based on data from the TIMMS released by the NCES in 1992, his research showed student scores remained about the same or slightly improved compared to the previous SISS study. Instead, Bracey focused attention on the NAEP; the data shows a consistent decline since 1975 in the gap between white and Hispanic scores. According to Bracey (1996), “the gap between the scores of black students and those of white students declined from 1975 to 1988, but increased in 1990 and again in 1992” (1996, p. 132). He found even more disturbing “that the gap remains large: black 17 year-olds score just below white 13 year-olds on the NAEP Mathematics test” (p. 132).

In a research study conducted by Oakes, Wells, Yonezawa, and Ray (1997) data was collected in 10 racially and social-economically mixed secondary schools that used ability grouping as a means of tracking students in a three-year longitudinal study. As a by-product of tracking, racially separate programs that

provided minority children with restricted educational opportunities and outcomes were created. Oakes' study revealed patterns that became apparent among the schools as they were detracked and students were placed in heterogeneous ability groupings.

One noted outcome of the study was an attitude towards the reallocation of resources such as time, teachers, materials and high achieving students and traditional ways of thinking about merit. Questions were raised as to which students deserved the best that schools had to offer. Stereotypical racial views remained salient, as did cultural biases.

Educators adopted the view that all students could achieve at very high levels and communicated that expectation through all school structures and programs. Use of new theories about the multidimensional nature of intelligence provided means for assessing achievement. Broader arrays of instructional strategies were employed to stimulate and motivate the heterogeneous grouping of students. On all but one campus, achievement-test scores remained steady or even improved.

Lessons learned from the study include the need for bottom-up reform efforts, along with top-down support structures already in place. The findings also suggested that when dealing with powerful race- and class-linked changes, the typical change barriers might blind the reform leadership to the needs to democratize the school experience. As a final note, due to pressure by parent organizations and community members with a strong sense of entitlement, not all programs were detracked, most notably the Advanced Placement programs. This

resulting situation allows again for the “white and wealthy” (p. 72) parents to maintain the ability to drive the ongoing curriculum differentiation that inevitably leads to unequal standards for students of different races, social class, and cultural background (Oakes, et al., 1997).

Often studies are cited to explain and to excuse the low achievement in schools that have a population affected by low socioeconomic status. According to Marzano, (McREL, 2001) Coleman et al. (1966) and Jencks et al. (1972) did not interpret the 10% variance in achievement as percentile gains in achievement. It is important to consider and identify the factors, which do provide gains in achievement for students in poorer schools over those in average schools. Specific instructional strategies and practices identified to be effective for achieving gains are part of the framework in the *Standards*.

Unfortunately, the standardized assessment used for these students is often not appropriate based on the type or quality of instruction they have been receiving. The ensuing failures have many long-range affects. High-stakes standardized assessment have consequences and punishments that result in motivation and self-esteem issues that are often overlooked in the need for data analysis and justification of standards. High drop out and failure rates are often the results. In a research study conducted by Popham (2001), standardized test items were analyzed for bias. It was found that achievement test items were biased towards affluent and middle-class families whose children grow up in a home environment rich in materials and experiences such as books, cable TV and the standard American vocabulary. Students from low-income families do not

have exposure or meaningful experiences such as these on which to rely to answer the questions correctly. Children's responses to test items are influenced by parental education levels and family income, factors over which the school has no control.

As pointed out by English (2000, p. 66), socioeconomic status does predict what a student's score will be on a standardized test, far more than the school's curricula or its size. That is the case, when the test is poorly aligned with the local curriculum, according to Fowler and Walberg (1991, cited in English 2000). Curriculum alignment is a reference to and the process used "for the 'match' or overlap between the content and format of the test, and the content and format of the curriculum" (English, 2000, p, 63). When they match or are aligned, there is a better chance for higher student achievement on the test.

Williams (1999), in a study of the challenges for the education of diverse populations, stated that because cultural diversity has not been valued, preparation of students from diverse populations thus far has not been successful because there is no recognition and respect for their individual human rights, and there is a need to enable all students to participate in and contribute to the growth of the nation and the world community. The means to meet these challenges are threefold:

1. Comprehend and accept the paradigm shift in the conceptualization of diversity supported by the new understandings in human development.
2. Centrally position these new understandings in reform proposals introduced to increase the learning success of a diverse student population.

3. Integrate available models and strategies to facilitate conversations in efforts to develop the abilities and attitudes necessary for current populations of diverse students to successfully participate in the global markets of the 21st century. (p. 91)

Williams' study further suggests that top-down mandates will not work unless there is a new accountability system or a new instructional device to insure compliance. Teachers bring varying degrees of knowledge and skill to address the education of culturally diverse students. If site-based decision-making is intended to improve academic outcomes for all children, then the issues of diversity and equity must be clarified, defined, and scheduled for discussion.

In research conducted by Darling-Hammond (1996), she suggests "allocating resources and time to the central task of classroom teaching and teachers learning, ...restructured schools have managed to create democratic learning communities that succeed in ways not previously thought possible with diverse groups of students" (cited in ASCD, 1999, p. 104). In a similar study, conclusions by Williams (1999) state

There is a large gap between research-based reform efforts demonstrating only improved learning outcomes for culturally, linguistically, and socioeconomically diverse students and efforts resulting in comparable results for all students. (p. 105)

Resources for programs and innovations to successfully engage diverse student populations were identified and evaluated by Williams (1999). Some of these include differentiation strategies, looping, the use of authentic work and fostering resiliency for students, educators, and schools.

Through a standards-based system of instruction, assessment,

curriculum, and teacher preparation, attention can be brought to the equity needs of under-served populations. With additional funding from state and federal agencies such as Title I, local school districts will be able to provide instructional resources, additional services, and manpower to meet the needs of their low achieving students (NRC, 1996).

Standardized Tests and Accountability for Standards-Based Instruction

As policy makers try to “count” curriculum using quantifiable means to provide education accountability, Darling-Hammond & Snyder (1992) argue that it encourages quantification of many aspects of education. There are numerous valuable educational experiences taking place in classrooms that cannot be quantified but standardized assessments are used to evaluate the extent to which goals for the educational system are met through these experiences. Not only can the hours spent in instruction be counted, other data that is readily available, such as test scores, can be precisely analyzed. “Ways of measuring curriculum are also useful in controlling it” (p. 58). Many states use the data to legislate expectations. According to Darling-Hammond and Snyder:

Metrics must be devised to meet these needs since a constant way of counting is necessary when comparisons are to be made or when standardization is sought. ...Their context and meaning may be meanwhile forgotten, if ever known, by users. Or regrettably, their meaning and appropriate inference may never have been given much attention by their creators to begin with. (p. 58)

As further described by Darling-Hammond & Snyder (1992), the political importance of these measures as educational indicators, to get public attention to specific concerns such as low test scores or comparison to national or international studies, has the potential to provide incentives to the system by

allocating rewards to schools and teachers and conferring or denying educational opportunities to students.

It is a fallacy to think students have shown competency in science based solely on the determination of one science assessment in fifth or eighth grade. Assessments should be summative, formative, educative, and evaluative to move students beyond a competency measure. According to Luft (1999), “these other forms of assessment provide feedback to the student while informing both students and teachers about the learning process, curriculum, and instruction” (p. 43). Summative assessments reveal what students know about concepts, formative assessments provide information about how students become better learners, evaluative assessments furnish teachers with knowledge about their science instruction, and educative assessments assist students in learning how to learn.

When norm-referenced tests are used, according to English (2000), it would be unethical to “teach to the test” (p. 77) if the purpose of the test is to assess a random variable within the population to be assessed, such as judging the quality of teaching as it is directly measured using student test scores. Teachers have no control over what content is being tested, so the best they could hope for would be a bell curve for their students’ achievement, with a 50% failure rate. If test items were taught, the scores would no longer be random. English states that teachers cannot be blamed for poor test scores if the teachers’ behaviors have no impact on learning because the content of the test is secured.

Glickman (2001) sees the movement for state accountability as a complex phenomenon that came in on the coat tails of the standards-based education movement. It has increasingly locked teachers and schools into focusing their teaching on high-stakes tests that reward or punish schools. The National Governor's Association advocated tight controls over educational standards and recommended that they be included as part of the Goals 2000 legislation.

Because standardized tests use the multiple - choice format to test basic skills, statewide assessments may have an effect different from the intended outcome. This testing format of standardized tests puts a premium on memorized and isolated facts in comparison to understanding of science concepts (CPRE, 1996). English called this phenomenon "deskilling" the teacher.

The simple fact is that, the more tests are used to calibrate the success of learning in schools (and, by inference, of teaching), the more curricular materials are developed to focus (and thus limit) the viable options teachers may select. The use of any materials, however, ultimately has a similar impact. (Apple, 1979, cited in English, 2000, p. 14)

The *Standards* stress classroom assessment as a critical component in the classroom instruction system, balanced with curriculum, in the selection of instructional materials and strategies to be used to teach the standards (NRC, 1999a). But, if standardized testing is implemented in circumstances where curriculum and assessment are not balanced or aligned, then classroom instruction may become severely distorted (AERA, 1996).

Reviewing and summarizing studies by other researchers, Glickman (2001) has proposed several alternatives to powerful special interest groups that want a standardized system at national, state, and district levels. He suggests

there are dangers in homogenizing our educational system; the diverse needs of students and communities cannot be met by a system that assesses achievement through standardized testing. The suggestions include rebelling openly (Schrag, 2000), providing an alternate accountability plan (Gallager, 2000), assisting students to find ways to become involved in other types of learning (Scheurich, 1998), and honoring diversity by viewing education through the eyes of another culture (DuBois, 1949/1970, as referenced by Glickman, 2001, p 50).

Eisner (1999) places performance assessment into a broad educational and social context by stating, "Performance assessment is a closer measure of our children's ability to achieve the aspirations we hold for them than are conventional forms of standardized testing" (p. 1). Instructional practices that include understanding context, making judgments, and opportunity to act in order to try out speculation are of critical importance. New kinds of assessment are needed to determine the levels of student achievement based on this kind of thinking.

Despite the lack of a single definition, performance assessment is aimed at moving away from testing practices that require students to select the single correct answer from an array of four or five distracters toward practice that requires the student to create evidence through performance that will enable assessors to make valid judgments about 'what they know and can do' in situations that matter. Performance assessment is the most important development in evaluation since the invention of short-answer tests and its extensive use during World War I. (Eisner, 1999, p.2)

Eisner goes on to state that performance assessments do not fulfill the desire to hold schools accountable, and that accountability is facilitated only if students

can be compared. He calls these assessment practices “temperature taking” (p. 3), which makes these “global comparisons... for world-class schools” more than treacherous. They make the public anxious about school productivity. He suggests the use of two different kinds of assessment, one that is standardized and “one that reveals the distinctive talents of individual students and the effects of school practice on their development” (p. 4).

English (2000) points out that when the alignment of the test and local curriculum is known, test data has a great deal of relevance to local educators and then becomes feedback. The data regarding actual learning of students can then be attached in meaningful ways to what teachers should and ultimately do teach their students. With the state mandated tests used, the value of the data obtained lies in the extent to which it the teacher aligns instruction with the tested curricular content.

Assessment of Student Achievement in Texas

Following the directions and vision of Goals 2000, the Texas Education Agency chose an accountability plan that uses a highly prescriptive and mandated means to gather data about student achievement over time. The TAAS Technical Digest states, “The goal of the assessment program in Texas is to measure student progress toward achieving academic excellence. The primary purpose of the state student assessment program is to provide an accurate measure of student achievement in the areas of reading, writing, mathematics, social studies, and science. The results are used as a gauge for institutional accountability” (TEA, 1999, p. 1).

The TAAS is a product of the evolution of accountability in Texas. The focus of assessment changed as the standards changed, from minimum skills to academic skills, testing higher-order thinking skills and problem-solving ability. “As required by statute, Texas assessed minimum basic skills with the first Texas Assessment of Basic Skills (TABS) tests, then the Texas Educational Assessment of Minimum Skills (TEAMS) examinations. In the fall of 1990, changes in state law required the implementation of a new criterion-referenced program, the Texas Assessment of Academic Skills (TAAS) tests” (p. 1).

According to the Technical Guide (TEA, 1999), in November 1995, the State Board of Education adopted new rules for student assessment that reflected the enactment of Senate Bill 1. These rules comprise the Texas Administrative Code, Chapter 101. The TEA holds districts, schools, and students accountable for the levels of student achievement through the use of the TAAS (TEC, 1997). Policy makers enacted state-mandated testing with the intention of tracking improved student achievement trends. Schools are evaluated on both the percentage of all students who pass the TAAS and on the percentage of low-income and minority students who pass the tests. Reporting test results can be beneficial in directing public attention to gross achievement disparities among schools or student groups and in the redirection of state education funds; the rewards for doing well and the sanctions for doing poorly are both implicit and explicit (CPRE, 2001).

The Technical Manual points out that the TAAS scores are only a “snapshot” (p. 2) of student progress and if used appropriately can provide a

valid indicator of student performance. TEA reports concerning TAAS data can be used for several purposes such as the evaluation of student scores for placement; of programs, resources, and staffing patterns; and of district and campus instruction. The scores are reported to parents and school professionals providing information about passing standards, remediation needs in content areas, specific skills that need further diagnosis, as well as performance in comparison with the peer groups (1999).

The final TAAS scores reflect several adjustments to the original raw score. The TAAS testing program provides two derived scores: scaled scores and Texas Learning Index (TLI). Scaling is associated with 70% of the raw score of the original 1994 test. It is calibrated into a factor called R. The purpose of the scaled score is to ensure that the passing standard is maintained at the same level of difficulty across administrations. "The value of the Rasch model (R) assumes that for each raw score point, there is only one ability estimate" (p. 38). The TAAS scaled score has a range of approximately 400-2400, with 1500 corresponding to the 70% of the items correct on the 1994 administration of the test, where the passing standards were set.

"The TLI is an index of a student's performance relative to the mean performance of the other students in the same grade" (p. 40). It was developed for longitudinal comparability. The purpose of the TLI was two fold: to provide an index to gauge student progress towards the exit level TAAS test required for graduation and for comparability between test administrations and between grades for accountability. The TLI is computed by turning the raw score into a T-

score, which is much like computing a z-score but it is re-anchored to have a mean of the exit level passing standard of 70 with a standard deviation of 15.

It is important to understand when comparing TLI for students that there would be little or no gain anticipated between grades where students are making typical progress. That is interpreted to mean the student is neither improving nor falling behind if there is little change in the TLI. Students who show an increase from one year to the next are gaining more than one year's growth, and if TLI shows a decrease then gains toward achievement on the exit level TAAS are at risk. Based on data provided by the TEA of the 180,000 students who took 8th Grade TAAS tests in spring of 1997, 94% also passed the exit level tests in 1999. "Whereas, the more than 51% who failed the 8th grade tests failed the exit level tests" (1999, p. 44). "As an indicator of student future success... school districts may find it best to begin remediation immediately for students who fail certain portions of the 8th grade TAAS so they are well prepared for the high school exit level tests" (p. 44). Because the science TAAS is not given at each grade level, the TLI is not used to report the results of the 8th grade or biology EOC results.

Professional Development for Improved Student Achievement

According to Asa Hilliard III, "Revolution, not reform, is required to release the power of teaching" (1996, p. 1). In a keynote address at the 1996 National Staff Development Council meeting she also stated that

There is a critical problem with the deep structure of staff development in its traditional form – it cannot produce teachers who are routinely successful. The true measure of our success lies in how we perform and how the performance produces significant

positive outcomes when we meet children of the poor and of minorities, and not merely with children of privilege. (p. 1)

When evaluating the current state of student achievement in schools with large student populations of low income and minority students, Hilliard (1997) points out that there is a need to consolidate what is known about successful staff development practices that lead to the highest student achievement, documenting these practices through visual media to share the varied experiences of good teaching and good teachers. Hilliard has identified four areas that focus on the imperatives in staff development design, which lead to student achievement. These include ensuring an appropriate level of valid teaching skills in general, validating empirically that present pedagogical and assessment practices are absolutely essential, preparing teachers in a straight forward manner, and ensuring that staff development is efficient, rapid, articulated, and documented.

As advocated by Sparks (2000), student achievement data should be the basis for decision making about the use of staff development funds. Gearing staff development initiatives to what students are learning and what is being assessed is the highest priority. When data is used to track improvements in student achievement, the instructional environment becomes non-threatening and the barriers created by fear and fatalism diminish. Sparks sites the Brazosport ISD, a school district south of Houston, as an example of where data-driven staff development is credited with student achievement gains. According to Powers (1999) as part of Brazosport's staff development, teachers would then implement

throughout the year new strategies to advance student performance related to each school's goal.

Schools that approach this task seriously should expect to see almost immediate changes in student performance, and then they should see those changes continue from month to month throughout the year. Joyce and Showers' research (1986) supports that measurable improvements in student achievement will occur when staff development focuses on specific elements identified by data as low performing and implements strategies targeting those elements. In other research by Joyce and Showers (1982), factors that affect the relationship between staff development and student achievement include the quality of training and the value of sustained follow-up activities such as coaching, action research, or focused study groups and the provision for information, multiple demonstrations, and regular practice with the skills.

Staff Development Standards

A seminal study for the National Commission on Teaching and America's Future entitled "What Matters Most: A Competent Teacher for Every Child" (Darling-Hammond, 1996) is considered a "blueprint for recruiting, supporting, and rewarding excellent educators" (p. 2). It reports the link between the education standards and student achievement is quality staff development, stating if "a caring, qualified teacher for every child is the most important ingredient in education reform, then it should no longer be the factor most frequently overlooked" (p. 2). The following factors are necessary to close the gap in student achievement: first, training in the knowledge and skills needed by

teachers to meet the needs of culturally diverse student populations; second, standards with expectations for student learning and for teaching; and third, equalization of funding for quality education for all students. “Clearly, if students are to achieve high standards, we can expect no less from teachers and other educators. Our highest priority must be to reach agreement on what teachers should know and be able to do in order to help students achieve” (1996, p. 6).

Darling-Hammond (1996) states that restructuring of the educational system as it presently stands will be required. To address this issue, goals for the nation must include the following: “All teachers will have access to high-quality professional development, and they will have regularly scheduled time for collegial work and planning” (p. 5) and “High-quality teaching will be the central investment of schools. Most education dollars will be spent on classroom teaching” (p. 5). There is a need to “reinvent teacher preparation and professional development” (p. 7) if these goals are to be met.

In research by SEDL reviewed by Morrissey, Cowan, Leo and Blair (1999), findings indicate educator involvement in professional learning communities (PLC) improves the teaching and learning process through the use of inquiry.

As a PLC, the faculty members ask themselves if what they are doing in the classroom is effective? If it is not, they must ask themselves, what do we need to do differently? Then the group must decide how they should change its practice so that the students benefit. (Hord, 1999, quoted in Morrissey et al., 1999, p. 8)

In the February 2000 newsletter for the National Staff Development Council (NSCD), Dennis Sparks stated that “those seeking to reform schools

exclusively through standards, rigorous tests, and strict accountability measures seem to believe that teachers and administrators already know what to do to improve student achievement but just need to be prodded harder to work” (p. 1). The NSCD refuted this notion and calls for spending at least 10% of a school district’s budget on professional development, and 25% of teacher’s time on collaborative learning and planning. Student achievement depends on “what teachers know about the subjects they teach and their ability to use a variety of methods to reach an increasingly diverse student body” (p. 1).

Howard (1996) of the Texas Staff Development Council stated, to accomplish school reform, staff development must simultaneously address “context, content, and process of staff development” (p. 1). The NSDC in its *Standards for Staff Development* also bases the conceptual framework of staff development on the dimensions of context, content, and process. All three must be considered in the planning and implementation of any professional development program (1995).

According to Guskey and Sparks (1996), quality is central to any staff development program. “Staff development’s influence on students is accomplished principally through its direct effect on teacher and administrator knowledge and practice” (p. 2). In the model proposed for studying the relationship between staff development and improvements in student learning, they identified “factors or components in the model that strongly affected the relationship and lie within a school’s sphere of influence” (p. 2). The factors were classified into content characteristics, process variables, and context

characteristics. According to Sparks and Guskey, content characteristics refer to the “what” of staff development, the understanding of specific academic disciplines as well as the pedagogical processes. Process variables refer to the “how.” Based on the research by Sparks and Loucks-Horsley (1989, cited in Guskey and Sparks, 1996, p. 2 of online text), process deals with the way the staff development activities are carried out as well as the quality of the training and the value of sustained follow up activities. Context characteristics refer to who, when, where, and why of staff development. It is the culture of the setting. According to Guskey and Sparks, context also involves the culture of the school district or school where the training takes place, as well as the culture of the school where the new learning will be implemented (1996).

When studying the models for staff development proposed by Guskey and Sparks (1996) the following were found: (1) the relationship between staff development and improvement in student learning is complex, it is not random or chaotic; (2) the results are measurable and can be documented; improvements in student learning can be explained; and (3) staff development needs to be viewed from a systems perspective and reformed in a systemic approach. It made clear the critical roles teachers, administrators, and parents play in the complex relationship between staff development and improvement in student achievement.

The success of any reform effort is hinged on the methods of its implementation and the changes that it causes. Understanding the assumptions about change was the basis for the research by the Research and Development

Center for Teacher Education (Hord, Rutherford, Huling-Austin and Hall, 1998) which led to the development of the Concerns-Based Adoption Model (CBAM) to be used as a tool for change facilitators. The model is based on the understandings that

- Change is a process, not an event.
- Individuals must be the focus not the program.
- Institutions will not change until its members change.
- The change process is extremely personal experience and how the individual perceives it will greatly affect the outcome.
- Individuals progress through various stages regarding their emotions and capabilities regarding the innovation.
- The availability of diagnostic/prescriptive models can enhance the individual's facilitation during staff development.
- People responsible for the change process must work in an adaptive systematic way where progress needs to be monitored constantly. (1998, pp. 5-6)

The components of the CBAM model include the Innovation Configuration, Stages of Concern, and Levels of Use. "The key to successful facilitation is to personalize one's interventions by focusing attention on the concerns of those engaged in the change process and accepting those concerns as legitimate reflections of changes in progress" (Hord et al. 1999, p. 90).

In a study conducted by Loucks-Horsley and Pratt (1993), using the CBAM model for the implementation of middle school science curriculum, the evaluator determined the following: there was a wide range of implementation in the curriculum; there was a difference in implementation between teachers who were trained in the curriculum and those who were not; there were teachers who were trained in the curriculum who demonstrated low levels of implementation, and others who were not trained and who were successful implementing it; and

finally some components of the program were easier to implement than others. The researchers felt there was evidence of a paradigm shift in the nature and quality of science instruction at the middle school studied based on the CBAM model.

Inquiry-Based Staff Development

Teachers need to know more than just the facts to be effective instructors; they need to understand the underlying concepts of the discipline if they are to implement the kind of instruction reflected in standards-based reform, according to St. John (1999). Teachers also need to know the processes of inquiry for effective instruction in standards-based reform; questioning, grouping, managing students, recording, facilitating discussions, and so on. In planning staff development, although process skills are important, they often become the only focus to the detriment of an understanding of the underlying concepts.

In an evaluation study of the Exploratorium's Institute for Inquiry, (IFI) which supports an inquiry-based staff development program, St. John (2000) stated that the institute is focused on practical work, assisting education leaders across the nation who are implementing elementary science education reforms in their home schools and districts. Through its professional development offerings and its curricular resources, IFI is seeking to help these leaders make inquiry a central feature of their science education programs (p. 1).

As the process that creates the ever-growing body of knowledge, inquiry is the process that underlies all science; it is the vehicle for making meaning of experiences and allowing individuals to become conversant with and confident in

“what they know – and equally important, what they do not know” (2000, p. 4). St.

John goes on to state

Student achievement in science depends, in part, on what students learn in classrooms. And what they learn in classrooms depends in part, on the nature and quality of instruction they encounter there. And the quality of that instruction is itself highly dependent upon multiple critical system components – such as the quality of the teacher, and the soundness of the curriculum, etc. In turn, the strength of these system components depend, in part, upon the degree to which there exists a local ‘improvement infrastructure’ that is capable of providing continuing resources and processes that can upgrade the quality and effectiveness of the key system components that are needed for good instruction. Hence, good inquiry based science education depends upon the existence and efficacy of the local district improvement infrastructure, and it is this infrastructure that the Institute for Inquiry seeks to support the IFI program. (St. John, 2000, pp 3-4)

St. John, (1999) in a preliminary study, states that professional development needs to strike a balance between “getting it going and getting it good” (p. 48). In a program for standards-based reform, it is important to realize that both are needed. For establishing a floor of practice for large numbers of teachers, it is best to focus on “getting it going” (p. 48) through materials-based workshops. For other teachers who express interest and willingness, more intensive and deeper learning experiences in inquiry and content are appropriate. Ultimately, the approach needs to be balanced, where there is attention paid to “getting it going” but also efforts made to help teachers “get it good,” to set the floor, but also to raise the ceiling. It is important is to realize that the two types of approaches are different, are appropriate to different teachers at different times, and require quite different types of efforts.

Loucks-Horsley et al. (1998) reviewed several professional learning

strategies and determined the strategy entitled *Immersion in Inquiry into Science and Mathematics* as effective, based on the stated purposes to develop awareness, build knowledge, and to encourage reflection about the process.

The teachers receive extensive experience in which they focus on learning science and are able to pursue content in-depth. As learners they participate fully in the generation of investigable questions, plan and conduct investigations that allow them to make meaning out of the inquiry activities, collect and organize data, make predictions, measure and graph, and gain a broader view of the science or mathematics concepts they are investigating. They do all this to experience and learn the inquiry process for themselves. (p. 50)

The review states the findings of the program to be beneficial, but the drawbacks include teachers with limited time and programs with limited resources. Another consideration for the success of the program is the timing of when it best fits into the learning sequence.

In an ongoing follow-up study by Martinello (1997) of graduates from an elementary teacher education program, which used inquiry as the basis for instruction, the teachers reported their experiences in learning and using inquiry-based instruction. Reportedly, 64.36% believed it helped them to become self-directed learners. Of the group of 216 students, 69.91% believed that the program helped them to understand how knowledge is discovered, invented, and created. Then, 75.46% perceived the program as helping them learn to see connections among ideas from different fields of study.

According to Martinello, the key to successful teacher use of inquiry and exploration is the ability to hone the queries that direct one's search and take varied perspectives on the problem. Teachers must learn how to inquire using

the language of questioning and logical reasoning and the methods of research (Tishman, Perkins, and Jay, 1995 – as cited by Martinello, 1997). She states, “Many teachers are not initiated into that culture through their preparation for teaching or during their employment as teachers” (p. 167). How are teachers to be implementing standards-based instruction using inquiry when they have been neither trained nor given time to hone their instructional skills in inquiry-based learning or instruction?

Professional development strategies that succeed in improving teaching share several features according to the research of Darling-Hammond (1995). “They tend to be grounded in participants’ questions, inquiry, and experimentation as well as profession-wide research. They are also collaborative, involving a sharing of knowledge among educators” (p. 13). The professional development must include discourse around problems of practice and be content-based, not generic. Concluding her 1995 study, she stated, “Throughout their careers, teachers need to know their subject matter content thoroughly, understand the learning process through a child’s eyes with sensitivity to backgrounds and diversity, and to have the opportunity to sharpen their practice against the grindstone of experience” (p. 17).

Waxman and Walbert (1999) viewed teaching through the constructivist perspective by examining several aspects of teaching, which include teacher constructed knowledge, teacher as manager of instruction, student constructed learning, and using technology as a tool for constructivist learning (p. 7). The first perspective is in determining the legitimacy of narrative inquiry as an educational

research method. Horenstein (1995, as adapted by Waxman and Walbert, 1999) concluded that through story research, knowledge about teachers' expertise can be constructed and be accessible. It provided perspectives on teaching and learning that cannot be found in other forms of inquiry.

Evertson and Randolph (1999, edited by Waxman and Walbert, 1999) examined classroom management, focusing on the sociocultural aspects of management. They concluded that when students are highly involved in determining when and how to participate in classrooms, instruction changes from teacher-centered to student-centered. The research also showed that classroom management practices cannot be separated from instructional practices. The more student-centered the learning, the expectations participants have of each other and the definitions of teacher and learning evolve in the setting.

Ken Tobin (1999, as edited by Waxman and Walbert, 1999) explored constructivism as a way of thinking about knowledge and coming to know. There has been a tendency to equate constructivism with certain activities and room arrangement rather than the construction of knowledge and the process of that construction. The most important elements of this perspective include learning through co-participation and the evaluation of understandings through portfolio assessment. Tobin used the studies of Lemke (1995, as cited in Tobin 1999) to discuss how discourse about the learning is a way to make sense of an experience. It is a social endeavor to use language to make meanings of an experience or situation and that conceptual understanding, as related to the evidence and experience, is constructivist learning.

Fraser (1999, as edited by Waxman and Walbert, 1999) addressed the need for research based on a student's perspective of the learning environment, showing that a student's perceptions consistently account for appreciable amounts of variance in student learning. Finally, Means and Olson (1999, as edited by Waxman and Walbert, 1999) focused on the way technology can be used as a tool to support student thinking and productivity. They focused on the constructivist view of learning – teaching basic skills within authentic contexts, modeling skillful thought processes, and providing opportunities for student collaboration to achieve intellectual accomplishment.

Staff Development and Student Achievement

Many professional teacher organizations have stated their views concerning staff development. In a position statement submitted by Texas Council of Elementary Science, to the Texas Board of Education (1998), the council primarily advocated the use of an annual Science TAAS test in at least one elementary grade, and further strongly recommended, among other items, professional development with “topics that include content, methodology, current research, as well as safety rules and practices, and opportunities” (1998, p. 21).

In a National Science Teachers Association (NSTA) Position Statement, science teachers recognize the value of staff development and collegial efforts to achieve “the most meaningful science learning experience for students.... Teacher training and professional development programs must promote and help teachers work toward coordination for student achievement and quality teacher retention” (2000, p 1).

According to Anderson (2000), executive director of the Texas Staff Development Council, "Teaching quality is directly linked to student achievement." Findings by the National Commission on Teaching and America's Future found in 1998 that "every \$500 invested in teacher training has a greater impact on student learning than a similar investment in any non-instructional area." (p. 1) As stated by Joellen Killian, in *Results*, the publication of the NSCD,

Holding high expectations for students leads to increased student achievement. Students must believe they can succeed before they take a risk. Teachers are the key to helping students believe in themselves. Rarely will students choose not to succeed. Schools and teachers must create a learning environment of possibilities for students by believing in all students, teaching them in a way to promote their success, and committing to their achievement. (p. 1)

Conclusion

The literature review in this chapter began with a brief history and overview of the current science education reform movement that led to standards-based science education at the national and local levels. Issues and research regarding equity and quality issues were reviewed and discussed. Effective Schools research and the best practices in science were reviewed to provide background for the study concerning the levels of student achievement and mastery of the state and national standards. Finally, in this chapter, review of research described the role of professional development in the improvement of student achievement and the importance of inquiry-based learning in the process, context, and content of science instruction and professional development. As a result of this literature review, a rationale for further study of the effects of inquiry-based professional development on student achievement as

measured by state mandated standardized test could be developed. Chapter 3 will address the research design, methods and instruments of the study.

CHAPTER 3

RESEARCH DESIGN

Introduction

In this chapter, the purpose of the study is reviewed, and the research problem is discussed. The independent grouping variables and dependent variables are outlined along with the corresponding research hypotheses and the rationale for the quasi-experimental design. A discussion of the sampling methods is followed by a description of the population and sample. Justification for selection of the assessment instruments is given based on validity and reliability findings. The procedure for the experimental treatment is explained along with the assumptions associated with each type of treatment. Discussion of the data collection procedures leads to a description of the methods for the analysis of data in chapter 4. A summary and conclusion of the research design close the chapter.

Purpose and Statement of the Problem

It was the purpose of this study to examine the academic achievement of students whose teachers participated in an inquiry-based professional development institute and the academic achievement of students whose teachers did not participate in an inquiry-based professional development institute to determine if a relationship exists. Simply stated, the problem of this study is to determine the relationship between the participation of 8th grade

science and biology teachers in an inquiry-based professional development institute and student achievement as measured on the state mandated achievement tests, the 8th grade science TAAS and the biology EOC exam for the 9th and 10th grade students.

Research Hypotheses

It was hypothesized that the 8th grade science TAAS scores for students of Inquiry Institute teachers had:

H₁: higher mean scale scores than students of non-Inquiry Institute teachers.

H₂: higher passing rates on objective 2 than students of non-Inquiry Institute teachers.

It was also hypothesized that 9th and 10th grade biology EOC exam scores for students of Inquiry Institute teachers had:

H₃: higher mean scale scores than students of non-Inquiry Institute teachers.

H₄: higher passing rates on objective 5 than students of non-Inquiry Institute teachers.

Finally, it was hypothesized that students in identified subgroups, when taught by Inquiry Institute teachers had:

H₅: higher TAAS and EOC objective mastery rates than when taught by non-Inquiry Institute teachers.

Population and Samples

This study was conducted in one suburban school district in north central Texas. The total population in 2001 for grades K – 12 totaled 13,584. The percentages of student subpopulations are shown in Table 1. The sub-population in the school district totals 7.8% Hispanic, 2.9% African-American, 3.7% Asian, and 6.4% low socioeconomic status. The sub-population percentages are generally smaller or within 0.3% difference of the total population among the students taking the 8th grade science TAAS and the biology EOC.

Table 1. Total Student Populations (TEA, 2002c)

Group	Population	Hispanic	African-American	Asian	Caucasian	Low SES
District	13584	7.8%	2.9%	3.7%	85.0%	6.4%
8th grade science	1029	6.4%	2.7%	3.8%	86.2%	5.5%
9 th – 10 th grade biology	894	5.9%	3.2%	3.9%	86.0%	3.2%

In this quasi-experimental study, there were a total of 42 teachers and two assistant principals from the eleven elementary, four middle, and two high school campuses in the district who attended the Inquiry Institute as professional development in addition to attending traditional professional development. The principals of each campus selected the teachers to attend the Institute, using varied rationale. Teachers were selected based on their availability to attend the institute, their ability to be away from their classrooms for four days of training, as well as the principals' interest in receiving the benefits of the program for their campuses. In the total were four 8th grade preparatory advanced placement

(PAP) science teachers including one who also teaches regular 8th grade science. Also in the total were three 9th - 10th grade biology teachers who also teach PAP biology.

Each of the eleven elementary campuses had representatives, one from the primary grades (K-2) and one from the intermediate grades (3-5); each secondary campus had one representative from each grade and/or science content area. Participants in the Institute taught students in all levels of science, from those in special education to advanced placement (AP).

This study examined several student populations based on the classes in which they were enrolled such as (1) students enrolled in 8th grade regular and PAP science, and (2) students enrolled in 9th and 10th grade regular and PAP biology, as well based on specific student attributes within populations. Table 2 shows each population divided into four groups by teacher: those taught by teachers who attended the Inquiry Institute and those who were not; and those taught by teachers at the regular level and those taught at the PAP level.

Table 2. Teachers of Student Populations in 8th Grade Science and 9th -10th Grade Biology.

Population	Level	Teachers attending	Teachers not attending
8 th grade science	Regular	1*	6
	PAP	4	1
9 th – 10 th grade biology	Regular	1*	9
	PAP	3	1

*Same teacher also teaches PAP and is included in PAP total.

The number of students in each population is shown in Table 3. Each is divided into four groups, based on teacher attendance at the Institute and the

Table 3. Student Populations Divided by Institute Participants and Non-Participants

Science content area	N for Inquiry Institute participants	N for Non-Inquiry Institute participants
8 th regular science	47	557
8 th PAP science	414	17
9 th -10 th grade biology	75	685
9 th -10 th grade PAP biology	217	144

level. For the study, high schools were numbered 1 and 2, and middle schools 3 and 4. Because a sample of convenience was used in this quasi-experimental design, the teachers in the 8th grade science sample, as shown in Table 4, were equated by selecting 8th grade science teachers from the same campus who have between seven and fifteen years of teaching experience. The teachers in the 9th – 10th grade biology sample were equated by selecting teachers with 14 or more years of teaching experience, and for the PAP biology teachers, those with equal experience teaching advanced placement biology. Using the guidelines addressed by Gall, Borg, and Gall (1996) the limitations of non-random assignment to experimental groups are partially overcome when equivalence between intact groups is utilized. Teachers could not be matched by campus due to the limited number of teachers with more than 14 years of experience or AP experience. The two high school campuses are of similar student populations. As shown in Table 5, the student bodies are between .02 and 6.9 percentage points on each item reported in the 2001 Texas Education Agency Accountability Data.

The student sample was selected from two of the four middle school campuses (campus numbers 3 and 4) whose 8th grade science teachers attended or did not attend the Inquiry Institute and from two high school campuses (campus numbers 1 and 2) whose biology teachers attended or did not attend the Inquiry Institute. The student sample was equated for covariance using the students' 7th grade ITBS science scores to eliminate the effects of science knowledge and skills gained prior to the 2001-2002 school year.

Table 4. Intact class groups used for quasi-experimental design

Code	Group	N	Treatment	Years	Certification	Degree
1	Experimental biology PAP	25	Inquiry Institute as staff development	15 10 AP	Biology	MA
2	Control biology PAP	28	Traditional staff development	25 10 AP	Biology	MA
3	Experimental biology PAP	28	Inquiry Institute as staff development	14	Chemistry, biology	BA
4	Control biology regular	36	Traditional staff development	16	Biology, physical education	BA
5	Experimental PAP 8 th grade science	21	Inquiry Institute as staff development	9	Biology, chemistry, earth science	BA
6	Control PAP 8 th grade science	17	Traditional staff development	9	Composite science	BA
7	Experimental regular 8 th grade	24	Inquiry Institute as staff development	7	Composite science	BA
8	Control regular 8 th grade	31	Traditional staff development	15	Elementary, Early childhood	BA

Table 5. Comparison of Two High School Campuses Showing Equity in Schools

	Campus #1	Campus #2	Difference
Student population	2111	2024	87
TAAS rating	Exemplary	Exemplary	None
Attendance rate	95.2%	95.7%	.02%
African-American population	2.5%	2.7%	.5%
Hispanic population	6.0%	5.9%	.1%
Caucasian population	87.3%	89.1%	2.2%
Low socioeconomic status population	2.3%	3.7%	1.4%
Drop-out rate	.02%	.06%	.04%
Taking SAT	87.4%	80.5%	6.9%
Above criterion on SAT	43.5%	37.9%	5.6%
Completing recommended HS program	72.1%	76.9%	4.8%

Generalizations to populations were made with caution. Special Education students were not included in the data collection reported by the Texas Education Agency. Special Education status was based on the recommendations of the Admission, Review and Dismissal (ARD) committee. The ARD committee determined that the TAAS and EOC exams were not the appropriate assessments of the Special Education students' abilities and the instruction received.

Care was taken to maintain the confidentiality of all subjects in this study. A numerical code was assigned to each campus, each teacher and all student data used. No direct contact was made with students and data were gathered after the testing and experimental treatment was complete. The Human Subjects Review Board approved the experimental design.

Identification of the Variables

The independent variable for the study was enrollment in an Inquiry Institute program for teacher professional development. The experimental group treatment was teacher participation in the Inquiry Institute. The control group treatment was teacher participation in a traditional professional development program. The effects of the independent variable were observed over several dependent variables including: scaled scores on the 8th grade science TAAS and biology EOC, passing rates on TAAS objective 2 and EOC objective 5, and mastery rates on TAAS and EOC objectives. Examination of scores was based on student attributes to determine whether observed relationships were correlational. These attributes included gender, gifted and talented identification (GT), socioeconomic status based application for free or reduced lunches (SES), and identification for at-risk of failure or possibly dropping out of school (at-risk). Percent scores of the 7th grade science subtest on the Iowa Test of Basic Skills (ITBS) were used as covariates.

Instrument

The TAAS and EOC are criterion-referenced measures that are part of the required academic assessment program for the state of Texas. The 8th grade science TAAS is currently being administered in the spring to all students in the 8th grade. Students in both PAP and regular level 8th grade science took the exam. The biology EOC examination is administered to all regular and PAP biology students upon completion of biology in 9th or 10th grade. Most biology students are either 9th or 10th graders, but a few are 11th graders repeating the

course due to failure the first time taken.

Exam items on the EOC and TAAS are aligned to testing objectives. Test objectives are overarching statements under which the Texas Essential Knowledge and Skills (TEKS) and student expectations are meaningfully grouped. Not all TEKS are tested on the TAAS, but all test objectives are tested with the same number of items on every test. There are five areas tested on the 8th grade science TAAS exam:

1. Scientific Inquiry, Inference, and Communication
2. Critical Thinking and Problem-Solving
3. Living Systems, Interdependency, and Genetic Change
4. Characteristics of the Universe, Matter and Energy
5. Earth Systems, Natural Events and Human Activity (TEA, 2002a)

In biology there are nine main areas covered by the objectives on the EOC.

1. Heredity and Biological Changes over Time
2. Structure and Functions of Organisms
3. Patterns and Processes in Living Systems
4. Ecology
5. Design and Conduct Biological Experiments
6. Acquire and Organize Scientific Data
7. Interpret and Communicate Scientific Data
8. Make Inferences and Communicate Scientific Data
9. Apply Science to Daily Life (TEA, 2002a)

Eighth grade science TAAS items in objective two and EOC items in objective five are based on the TEKS and *National Science Education Standards* that specifically address the use of inquiry and the inquiry processes for problem solving and lab or field investigations. Student scores on these two objectives could be indicative of implementation of the inquiry-based professional development during classroom instruction.

The TAAS and EOC are formatted as multiple-choice tests. The format of

both tests is standardized and consistent across content areas. According to Sevenair and Burket (2001),

Multiple-choice questions have their own terminology. The question itself is the *stem*, and the alternative possible answers are *foils*. The entire question is an *item*. Incorrect foils are *distracters*; the correct response is the *key*. *Options* are a set of possible answers referred to in the foils. The *difficulty level* generally refers to the fraction of the tested population that answered an item incorrectly. An item's *discrimination* is a correlation between the students' performance on individual items and overall achievement on the exam (p. 6).

There are several significant strengths associated with multiple-choice tests, such as coverage of a wide range of cognitive domains ranging from rote recall to evaluations, conclusions, and judging evidence. Because multiple-choice items minimize writing requirements, more content material can be covered on the assessment. Feedback from disaggregated test data is easily calculated based on the analysis of answer choices selected and the comparison of data among students and from year to year. There are some limitations as well, according to Sevenair and Burket (2001). Real-world cognition differs from that which is tested on a multiple-choice test. It is a different cognitive process to propose a solution to a given problem when the solution must be selected from a set of alternatives than when a student must develop possible solutions without assistance.

Performance standards for the TAAS and EOC are based on students answering correctly a specified number of items for each objective. For each administration, the number of items needed to meet minimum expectations may shift in order to maintain equivalent passing standards (TEA, 2002b). On the 8th

grade science TAAS for objective 2, six out-of eight questions must be answered correctly for mastery expectations. Overall, 70% or 30 out-of 40 items must be answered correctly for a student to meet minimum expectations and pass the test. It is possible for students to have high scores, yet not master all the objectives, based on high achievement in some objectives and low achievement in others. To gain academic recognition, 95% or 38 out-of 40 items must be answered correctly. For each administration of the test, there are field test items that are not scored.

On the biology EOC, the raw score for meeting minimum expectations is 69% or 29 out-of 42 questions answered correctly. Objectives 1, 4, and 5 require 80% or five out-of six correct for mastery. Objectives 2, 3, 6, 7, 8, and 9 require 75% or three out-of four. To show mastery of all objectives, 79% or 33 out-of 42 items must be correct. To gain academic recognition, 95% or 40 out-of 42 must be met (TEA, 2002b). The results of the TAAS and EOC for the state are released by the Texas Education Agency as part of the district accountability ratings. The 2002 testing results based on these requirements (see Table 6) are higher overall for the district sample groups than for the state as a whole.

Table 6. Sample and State Percentages for TAAS and EOC

		Met Minimums Standards	Mastered all Objectives	Passed Objective 2 or 5
TAAS	Sample	99	64	89
	State	93	38	72
EOC	Sample	92	36	85
	State	68	15	44

According to the Dana Center (2002a), the goal of assessment for standards-based science education is to ensure that “the picture of student

learning is a complete dynamic view rather than a one time snapshot that may or may not show the actual amount of student learning. All stakeholders, especially students, should understand how to interpret the multiple assessments in order to understand where students are, how far they have come, and how they are advancing” (p. 2). The Dana Center advocates using a variety of assessment methods, both standardized and non-standardized. Each is appropriate when the selection is based on the purpose of the assessment, the emphasis of the test, the ability to document progress, the value of the information collected, and the reliability of the scoring.

The reliability of the TAAS is documented in the TAAS Technical Digest (1999) which states, “The test reliability is based on internal consistency measures, in particular Kuder-Richardson Formula 20 (KR-20).” “Most KR-20 reliabilities are in the high .80 to low .90 ranges. Reliability data for the TAAS for spring 1997 administration to 214,000 students show the internal consistency estimates for math, reading, and writing were 0.934, 0.878, and 0.838 respectively, and the standard errors of measurement were 2.876, 2.352 and 2.195” (Kellow and Willson, 2001, p. 12). Kellow and Willson (2001) found fault with TEA for using internal consistency for the reliability estimates. They state that more confidence should be placed in test-retest or alternate-forms reliability estimates. TEA argues that these reliability estimates are not possible since no student takes two versions of the same content area test. By recalculating “the results using test-retest correlations on student data for several large districts, math, reading, and writing subtests reliabilities were found to be: 0.643, 0.536,

and 0.555 respectively” (p. 12). Based on these findings and using calculations on alternate form reliability, Kellow and Willson (2001) suggest that standard errors of measurement may be on the order of 20-40% greater than that estimated by the TEA Technical Digest.

Establishing content validity is a process that involves collecting evidence to support the assertion that the test score implies knowledge and understanding of the Texas required curriculum (TEA, 1999). The TAAS review process, since 1990, has had over 7,000 Texas educators participate in committees to review new and field-tested items, verifying the alignment of the test items with the test objectives and test measurement specifications to ensure that the TAAS and EOC items measure appropriate content, and thus have content validity.

Based on recent legislative revisions of the Texas Education Code, the 8th grade science TAAS and the biology EOC are discontinued after the 2002 administration. A newly revised test, the Texas Assessment of Knowledge and Skills (TAKS), is to be benchmarked in 2002-2003 and used as the accountability system beginning in the 2003-2004 school year.

Procedures for Experimental Treatments

The treatment for this study was teacher participation in the Inquiry Institute. The Institute was based on the model of staff development developed by the Exploratorium in San Francisco, CA. The Exploratorium is a Science Museum, sponsored in part by the National Science Foundation, dedicated to the exploration of the natural world through hands-on interaction with scientific phenomena. According to the Exploratorium (2002), participants in the Inquiry

Institute “explore scientific topics while staff model and teach an inquiry approach to learning that can be transferred to any topic or classroom. Participants also explore natural phenomena through scientific investigation, careful observation, asking questions, probing for answers, and sharing discoveries” (p. 1). The Inquiry Institute, through a series of orchestrated activities, journals, and group discussions, develops in participants an understanding of underlying scientific concepts and the inquiry process of learning.

The Institute conducted for this study consisted of four, six-hour sessions that were spaced throughout the school year. The 42 participants met in September, November, February, and April. The instructional focus of the Institute was inquiry, which provided time for questioning, experimentation, discussion, and reflection. The content of the Institute was the use of inquiry to facilitate instruction of standards-based practices using the scientific method, tops, ice balloons, and shadows. The process of the Institute was inquiry. Teachers not only used inquiry in experimental investigations, but also used inquiry to discover the content and processes that provide a challenging yet rewarding experience to students.

The agenda for the four days of the Inquiry Institute was as follows:

- Day 1: Three Kinds of Science – Not all hands-on experiences are inquiry
Process Circus – Looking at the processes of inquiry
- Day 2: Ice Balloons – Developing investigable questions about ice
Introduction to Inquiry – A shadows based inquiry investigation
- Day 3: Shadow Investigations – Experimenting to answer inquiry-based questions, reflection, and common understandings

Day 4: Subtle Shifts – Moving into inquiry from cookbook labs;
Redesigning current practices to include inquiry; and looking at
assessment data to focus on the needs of underserved students

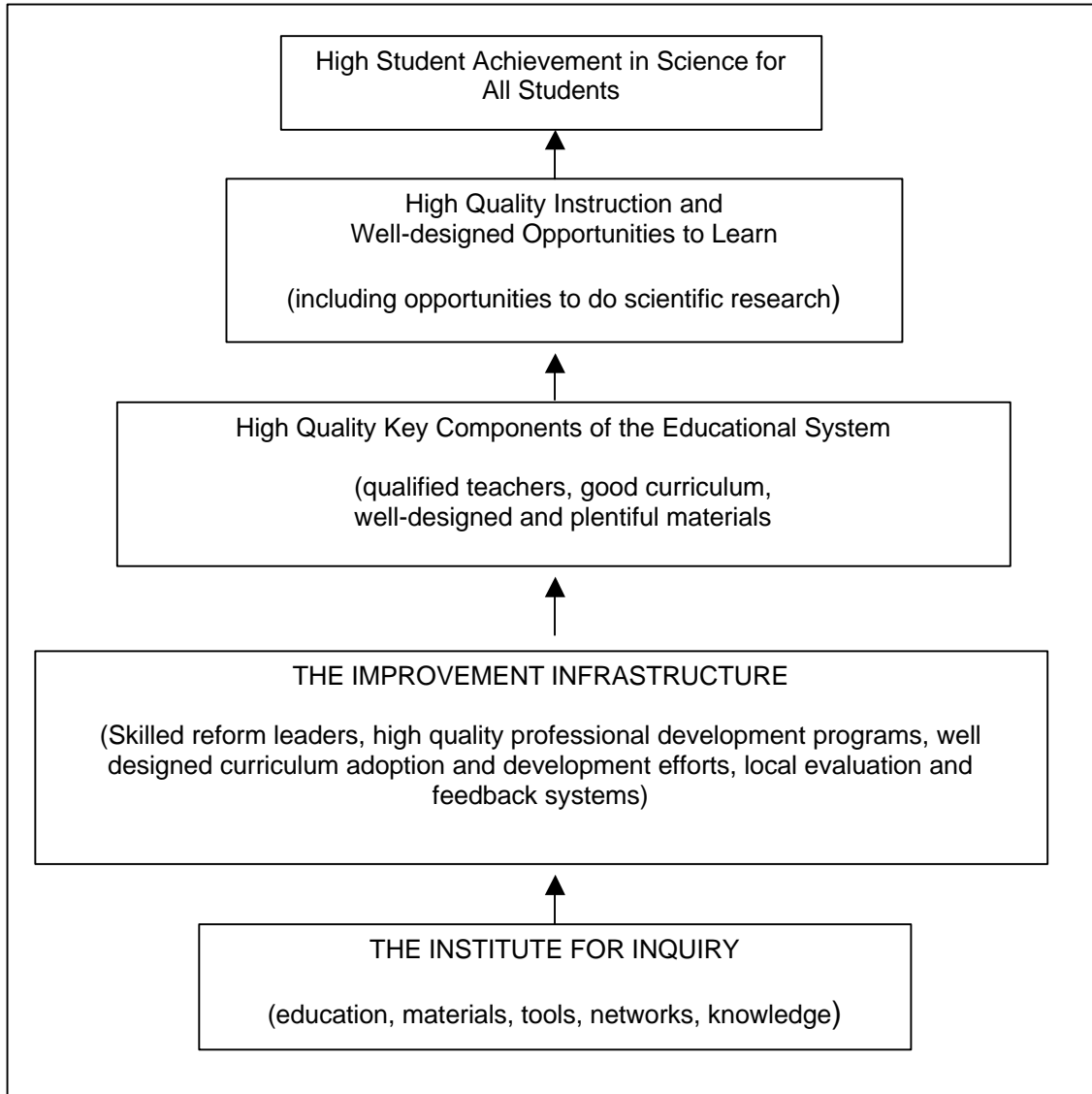


Figure 1 – Student achievement in science depends on the infrastructure that can support and sustain teacher knowledge and strategies of instruction (St. John, 2000, p. 3).

Sessions met from 8:00 AM until 4:00 PM daily, with a break for lunch. As addressed earlier, (see Figure 1) the Institute was grounded in the standards-based curriculum and standards-based assessment supported by the *Benchmarks, Standards, and TEKS* as well as the sound instructional practices

recommended by the effective schools reform movement.

All pairs of teachers attended 18 hours of traditional science professional development during the 2001-2002 school year, which included a six-hour curriculum update and lab strategy session. The Inquiry Institute was offered twice a year as part of the professional development program of the school district and could be selected to partially fulfill the professional growth requirements for the district.

Assumptions

Although all students are enrolled in eighth grade science or biology, it is not assumed they are all of equal ability levels. It is assumed that they are all able to function successfully in the class. It is assumed that matched and stratified selection of the experimental and control students groups from like-classrooms/schools will ensure that the study's results can be attributed to the independent variable, not to the initial differences in experimental versus control students. The ITBS 7th grade science subtest scores were used as a covariate to control these differences in the sample data.

Although all teachers in the study are certified to teach middle school or high school science, not all are of equal level of experience, tenure, or motivation. Teachers selected for the study were matched and equated based on these factors. It is assumed that matching the teachers on degrees earned and years of experience to within 5 years, will ensure that the study's results can be attributed to the independent variable, not to the teacher factor. It is also assumed that teachers who participated in the Institute will use what they have

learned when delivering instruction to their students.

It was assumed that students assigned to PAP or regular biology and PAP or regular eighth grade science classes were originally assigned to the teachers and class periods based on achievement levels to qualify for PAP enrollment. To avoid bias caused by non-probability sampling, the student groups were being equated for covariance based on prior 7th grade ITBS science scores, teachers' level of experience, and teachers' participation in the Inquiry Institute. Intact groups of students in classes, not individual students, were selected.

It was also assumed that instruction in eighth grade science and biology was based on the school district curriculum, which is based on the TEKS. It was also assumed that the TAAS is a valid and reliable assessment of the TEKS.

It was assumed that (1) student population samples selected from eighth grade science and the biology classes received equitable content exposure to the TEKS-based science and biology curriculum and (2) teachers used the allotted conference-preparation time to prepare for lab activities, which made up 40% of instructional time as mandated by the TEA in the Texas Education Code (TEC, 1997, Subchapter C, 112.41, p C-1).

Procedure for Data Gathering

Participation in the Inquiry Institute was documented through the district staff development records. The 2001-2002 dates for four sessions of the Inquiry Institute were September 25, December 11, February 7, and April 23. Teacher certification, degrees earned, and years of experience were available by a request for records from the district's human resources division. Student sample

data were gathered from several sources. Seventh grade ITBS scores for the current 8th, 9th, and 10th graders in the sample groups were requested from each middle school counselor's office and coded by student number and current teacher number. The window for administration of biology EOC was May 12 - 13, 2002. The results of the EOC were released by the Texas Education Agency (TEA) and available through the district's director of curriculum and assessment office on August 1, 2002. The 8th grade science TAAS test was administered statewide on April 19, 2002. The results for the TAAS were sent to the district by May 15, 2002. Through the PEIMS record keeping system, district and campus test data were available for access by district administration and teachers by August 1, 2002.

Summary

It was the purpose of this study to examine the standardized test performance of students whose teachers participated in an inquiry-based professional development institute and the standardized test performance of students whose teachers did not participate in an inquiry-based professional development institute. This study attempted to determine if a relationship existed. The focus of this study was to determine the relationship between the participation of 8th grade science and biology teachers in an inquiry-based professional development institute and student achievement as measured on the state mandated achievement tests, the 8th grade science TAAS and the biology EOC exam for the 9th and 10th grade students enrolled in biology.

Current reform efforts have brought science education to the attention of

the general public. Research is needed to investigate whether or not various strategies of professional development have an impact on student achievement. While investigating the effects of inquiry-based instruction as learned through inquiry-based professional development, the academic achievement of students in 8th grade science and 9th-10th grade biology were statistically analyzed and discussed extensively in the next chapter.

Instruction in the TEKS and the *National Science Education Standards*, as part of the standards-based curriculum, is only as effective as the instructional strategies and best practices that are used. Inquiry-based instruction is one of these highly effective best practices. If teachers elect to use instructional strategies that are inquiry-based, the process, content, and context of instruction and high levels of student achievement may become linked.

Conclusion

In this chapter, the purpose of the study was reviewed and the research design discussed. The population and sample were described, and the independent and dependent variables were identified. The sampling method was discussed, and the justification for selection of the assessment instrument was given based on validity and reliability findings. Finally, the experimental treatment procedure was outlined and the procedure for the collection of the data was reported. Chapter 4 will contain the analysis of the collected data and the hypothesis testing needed to determine statistical significance of the differences in the findings.

CHAPTER 4

DATA ANALYSIS

Introduction

Data analysis includes the organization and coding of the data for each analysis technique. Once the data are organized and coded, expected versus actual sample sizes are studied. Caution needs to be exercised in drawing conclusions from results based on small sample data sets. In this chapter, there is a systematic study of the data collected and its subsequent analysis. Conclusions are drawn based on meeting the assumptions of the methods of analysis, adequate sample sizes, prior research results, and statistical significance testing for the null hypothesis.

Collection and Coding of Data

The data for the study were gathered from existing student records at the school district level and at the campus level. The collected data were organized into a spreadsheet format then transferred into the SPSS data analysis program. Categorical student data such as gender, socioeconomic status, gifted and talented identification, and at-risk-of-failure status were coded and analyzed separately from teacher data. Categorical teacher data were also coded and included in the descriptive statistics to ensure that sampling procedures were adequate and based on assumptions for each method of data analysis.

Descriptive Statistics

All students enrolled in 8th grade science and all students enrolled in 9th grade biology at campuses 1 and 2 were the student populations from which the samples were taken. Tables 7 and 8 show the descriptive statistics for teachers and students in the study.

Table 7. Frequency Table for Teacher Attributes

Attributes/Teacher	1	2	3	4	5	6	7	8
Participant	Yes	No	Yes	No	Yes	No	Yes	No
Grade taught	9	9	9	9	8	8	8	8
Degree earned	MA	MA	BA	BA	BA	BA	BA	BA
Teaching experience (years)	15	25	14	16	9	9	7	15
N of students with TAAS / EOC scores	27	28	28	36	21	17	24	31
N of students with ITBS scores	24	20	23	21	20	17	19	21

Note. N for students in actual study is reduced based on availability of both TAAS/EOC scores and ITBS covariate scores

Table 8. Frequency Table for Student Attributes

Attribute		N TAAS*	%TAAS	N EOC**	%EOC
Gender	Male	41	44.1	41	34.5
	Female	52	55.9	78	65.5
Ethnicity	Native American	2	2.2	2	1.7
	Asian	5	5.4	3	2.5
	Hispanic	10	10.8	11	9.2
	Caucasian/other	75	80.6	103	86.6
Gifted and Talented		45	48.4	85	71.4
Requires Student Support Services		7	7.5	3	2.5
Free and reduced lunch qualified		3	3.2	2	1.7
At risk		11	11.8	13	10.9
Vocational program		39	41.9	31	26.1

* Excludes 27 students who lack ITBS scores, which were used for analyses.

** Excludes 30 students who lack ITBS scores, which were used in analyses.

Methods of Analysis

Based on the discrete and continuous format of the data collected, several techniques were used. A correlation analysis to test for a relationship between the TAAS and EOC scale scores and the ITBS percentage scores was conducted. Because the TAAS and EOC scale scores and ITBS data are continuous, analysis of covariance was selected to determine the differences of the means of these three tests, having controlled for ITBS performance. To determine the cause for differences in the mean scores of the TAAS and EOC, grouping variables such as gifted and talented identification (GT), socioeconomic status (SES), and at-risk data were analyzed. Categorical dependent variables such as these call for the use of logistic regression because of the binary nature of the dependent variables studied. The chi square test, based on cross tabulation, was also used when the data was nominal or categorical, such as gender or pass/fail. It was used for binary dependent and independent variables to predict the odds ratio for the TAAS and EOC objectives that specifically assess critical thinking, lab procedures, and problem-solving strategies.

Correlations

To establish the magnitude of a relationship between the ITBS 7th grade science subtest percentage score and the TAAS and EOC scale scores, a correlation analysis was conducted. The only student records included in the correlation analysis were those of students who had both TAAS and ITBS scores and students who had both EOC and ITBS scores. The proportion of the individual differences in the mean TAAS scale score compared to the mean ITBS

percentage score is reported as the Pearson product-moment correlation coefficient or Pearson r . Based on the Rule of Thumb used by Hinkle, Wiersma, and Jurs (1998), when the r is between .50 and .70 there is a moderate positive correlation, between .30 and .50 there is low correlation and when between .00 and .30 there is little if any correlation.

Students' scores on the ITBS scale were positively correlated with their TAAS scores ($r = .54, p < .01$). Likewise, students' ITBS scale scores were positively correlated with their EOC scores ($r = .57, p < .01$). For the TAAS and EOC, there was minimal negative correlation of the scale scores ($r = -.04, p > .05$; see Table 9). The correlation shows that there is a moderate relationship between the ITBS and the TAAS as well as a relationship between the ITBS and the EOC. Because a relationship exists between the two exams and the ITBS, the ITBS 7th grade science subtest percentage score was selected as the covariate to eliminate the effect of prior science knowledge in the analysis of variance for the TAAS and EOC scale scores. There was not a relationship between the TAAS used in 8th grade science and the EOC used in 9th grade biology. It would be inappropriate to use the TAAS scale scores as a covariate in the analysis of the EOC scores.

Table 9. Correlation Between TAAS and EOC Scale Scores and the ITBS Scores

	TAAS	ITBS	EOC
TAAS (N)	---		
ITBS (N)	$r = .54^* (n = 76)$	---	
EOC (N)	$r = -.04 (n = 89)$	$r = .57^* (n = 88)$	---

* $p < .05$ (2-tailed)

Analysis of Covariance

This study used intact groups of students that could not be randomly assigned to treatment groups. The ANCOVA analysis was selected to adjust for preexisting differences in science knowledge, as identified by ITBS scores, in each group prior to receiving instruction from Inquiry Institute participants or non-participants. The null hypothesis for the ANCOVA testing the continuous dependent variable of EOC scale score was $H_0: \mu_1 = \mu_2$, which states the mean of the EOC scale score when adjusted by the mean of the ITBS scores would be equal for participants and non-participants. The Levene test for the ANCOVA was robust, and the assumption of homogeneity of variance was met. The hypothesis of equal variance was not rejected. The calculated value of F for the Levene test was not statistically significant, with the p value = .98.

Although EOC scale scores were higher for students who received instruction from a participant than for students who received instruction from a teacher who was not a participant, the effect was negated when the influence of ITBS scores was controlled. (See Tables 10 and 11.)

Table 10. Mean EOC Scores for Students in Participants' Classes and Students in Non-Participants' Classes

Institute participant	M	SD	n
Non-participant	1807.86 ^a	173.1	42
Participant	1878.91 ^b	157.5	46
Total	1845.00	168.0	88

Note. Means with different superscripts differ at $p < .05$

Table 11. Analysis of Variance for Biology EOC Scale Scores Adjusted by ITBS Scores

Source	<i>df</i>	<i>F</i>	η^2	<i>p</i>
ITBS (Covariate)	1	38.33*	.31	.00
Participation (Between)	1	1.61	.02	.21
S within-group error	85	(19007.77)		

Note. Variable enclosed in parentheses represents mean square errors. S = students. * $p < .05$

When studying the results of the ANCOVA, the statistical value as well as the practical value of the study must be established. Eta-squared was used to determine the effect size or practical significance of the treatment. According to Hinkle et al. (1998, p. 568), the square of the eta coefficient (η^2) is the portion of the variance in the dependent variable that can be attributed to the variance in the independent variable. Pederson (2002) assigned values to the relative effect sizes. An effect size of .01 is a small effect size but would not be considered trivial. A medium effect size of .059 would be visible to the naked eye of a careful observer. A large effect size of .138 would be the same amount above a medium effect size as a small effect size is below a medium effect size. The ITBS test score covariate was used to increase the precision of the study by reducing error variance caused by the confounding effects of preexisting science knowledge.

Based on the results of the ANCOVA for the EOC scale scores adjusted by the ITBS scores, the calculated value of F for the EOC scores did not exceed the F_{cv} and was not statistically significant at $p > .05$. There was a small effect size for participation with $\eta^2 = .02$. ITBS scores had a large variance-accounted-

for effect size with $\eta^2 = .31$. The calculated value of F for ITBS scores as a covariate exceeded the F_{cv} and was statistically significant at $p < .01$. The null hypothesis which stated that the means for the two participant groups would be the same when the effect of ITBS scores were controlled was not rejected because there were no differences among the adjusted means for the groups.

A two-way ANCOVA for the TAAS 8th grade science scale score was used to test the null hypothesis $H_0: \mu_1 = \mu_2$, which states the mean of the TAAS scale score adjusted by the mean for the ITBS would be equal for participant and non-participant groups. Based on the Levene test, the ANCOVA was robust (see Table 12). The assumption of homogeneity of variance was met for the TAAS one-way ANCOVA, and the null for the Levene test was not rejected. The calculated value $F(1.68)$ did not exceed $F_{cv}(4.00)$ with $p > .05$.

Table 12. Mean TAAS Scores for Students in Participants' Classes and Students in Non-Participants' Classes.

Institute participant	M	SD	n
Non-participant	1715.53	87.1	38
Participant	1725.26	112.7	38
Total	1720.39	100.1	76

Note. Means were not statistically significant.

Based on the results of the ANCOVA for the TAAS scale scores, adjusted for the ITBS scores, the null hypothesis was not rejected. Students who were in participants' classes did not perform significantly better than students who were not in participants classes when the ITBS scores were controlled for (see Table 13). A significant effect for ITBS was observed with $F(1,73) = 15.27$, $p < .01$. The

calculated value of F for the ITBS scores exceeded the F_{cv} and was statistically significant at $p < .001$. There was a large effect size with $\eta^2 = .29$. The calculated value of F for the main effect of participation did not exceed the F_{cv} and was not statistically significant at $p > .05$. There was no effect size for participation.

Although students who were instructed by teachers who attended the inquiry institute performed better on the EOC and TAAS tests than students who were instructed by teachers who had not received the training, the overall results of the ANCOVA's showed that the average scale scores for the biology EOC exam and the TAAS 8th grade science tests did not vary to a statistically significant level when the effect of the ITBS score were controlled. The η^2 indicates a large effect size for the ITBS on TAAS (.29) and EOC (.31) exam scale scores. It appears that students who do well on the 7th grade ITBS science subtest will also do well on the TAAS exam in 8th grade science and the EOC 9th grade biology. Teacher participation in either the institute or in traditional staff development does not have a statistically significant or practically significant effect on EOC or TAAS scores.

Table 13. Analysis of Variance for TAAS 8th Grade Science Scale Scores Adjusted by ITBS Scores

Source	<i>df</i>	<i>F</i>	η^2	<i>P</i>
ITBS (Covariate)	1	15.27	.29	.00*
Participation (Between)	1	30.29	.00	.72
S within-group error	73	(7263.26)		

Note. Variable enclosed in parentheses represents mean square errors. S = students. * $p < .05$

To be certain that teacher factors such as degree earned, years of experience, and level of instruction were minimized by the selection of teachers from populations with equal attributes, an ANCOVA was utilized. A two-way analysis of covariance, testing for the between-subjects effects of participation and teacher factors, was not found to be statistically significant at $p > .05$ when analyzed for both the TAAS and EOC scale scores that were adjusted by the ITBS covariate score.

Student factors, such as gender as well as gifted and talented status, were also studied in two 2-way analyses of covariance using gender of the students and staff development type (participation and non-participation) as independent variables. The first ANCOVA tested for the effects on TAAS scores, whereas the second ANCOVA tested for the effects on EOC scale scores, both using ITBS scores as a covariate. The 2 X 2 ANCOVA's were not statistically significant at $p > .05$ and had no gender effect for the TAAS ($p = .34$) or EOC ($p = .55$) scale scores. The 2 X 2 ANCOVA interaction effects for gifted and talented X participation on the TAAS scale scores were, however, statistically significant at $p < .01$ (see Table 14) and yielded a large effect size ($\eta^2 = .15$). For the EOC scale score, the interaction effect for GT was not statistically significant at $p > .05$ (see Table 15).

Table 14. Two-way ANCOVA for TAAS Scale Scores using ITBS Covariate for Participation X GT

Institute participant	Gifted and Talented	Mean	SD	N
Non-participant	Gifted and Talented	1748.00	81.50	25
	Not Gifted and Talented	1653.08	60.74	13
	Total	1715.53	87.11	38
Participant	Gifted and Talented	1798.24	96.00	17
	Not Gifted and Talented	1666.19	88.96	21
	Total	1725.26	112.65	38
Total	Gifted and Talented	1768.33 ^a	90.04	42
	Not Gifted and Talented	1661.18 ^b	78.62	34
	Total	1720.39	100.14	76

Note. Means with different superscripts differ significantly at $p < .01$.

Table 15. Two-way ANCOVA for EOC Scale Scores using ITBS Covariate for Participation x GT

Institute participant	Gifted and Talented	Mean	S D	N
Non-participant	Gifted and Talented	1858.52	127.69	27
	Not Gifted and Talented	1716.67	208.83	15
	Total	1807.86	173.10	42
Participant	Gifted and Talented	1873.18	151.05	44
	Not Gifted and Talented	2005.00	19.00	2
	Total	1878.91	157.46	46
Total	Gifted and Talented	1867.61	141.86	71
	Not Gifted and Talented	1750.59	231.64	17
	Total	1845.00	167.98	88

Note. No means differed significantly at $p < .05$.

There are inherent problems associated with interpreting the results of statistical analyses when using criterion-referenced tests, such as the TAAS or Biology EOC. One such problem could possibly be the ceiling effect. According to Gall, Borg, and Gall (1996), “a ceiling effect occurs when the range of difficulty of the test items is limited, and therefore scores at the higher end of the possible score continuum are artificially restricted” (p. 533). The apparent minimal gain of

GT students could possibly be misinterpreted as a loss in progress towards student achievement when taught by participating teachers. With rates of 92% for the EOC and 99% for the TAAS the number of students passing at minimum expectations could be an indication of the low levels of difficulty on the state mandated exams. The school district's passing rates for students meeting minimum expectations were 99% for the TAAS and 92% for the EOC; the mean scale scores were 1710 for the TAAS and 1759 for the EOC.

Statistically significant differences were not observable between the two 2 X 2 (participation X SES) and (participation X at-risk) ANCOVA's for TAAS and EOC scale scores adjusted by the ITBS covariate. Perhaps this was due to the small cell sizes, unidentified students with these attributes, or students who could have been categorized incorrectly. An inherent problem associated with interpreting the results of this ANCOVA could be confounding variables in the identification of students as at-risk or low SES students based on their qualifying for free or reduced lunch. There are many students with low SES who will not apply for the free and reduced lunch program for many reasons. One might possibly be because of the stigma associated with it. As a result, there are many unidentified low SES or at-risk students in each group, which may confound the results. Although part of the proposal for the study, it was decided to delete the SES category from further consideration because it was a source of variance that was unaccounted for.

Chi Square

The chi square test of equal proportions was set up as cross tabulations. This nonparametric test was used to determine whether a relationship existed between teacher participation in the institute and mastery of all objectives by students taking the TAAS or EOC. Also studied was the relationship between participation and passing EOC objective 5 or TAAS objective 2. Chi square tests were run without controlling for other factors. The null hypothesis for the chi square tests stated the difference between the observed and expected proportion for sample one would be equal and in the same direction as the difference between the observed and expected proportion for sample two. The logic for this analysis is modeled in figure 2. The two dependent variables were pass/fail for mastery of all objectives and pass/fail for the objectives measuring inquiry-based skills.

Figure 2. Model for predicting whether a student would pass/fail test objectives and for predicting mastery/non-mastery of all objectives based on teacher professional development participation

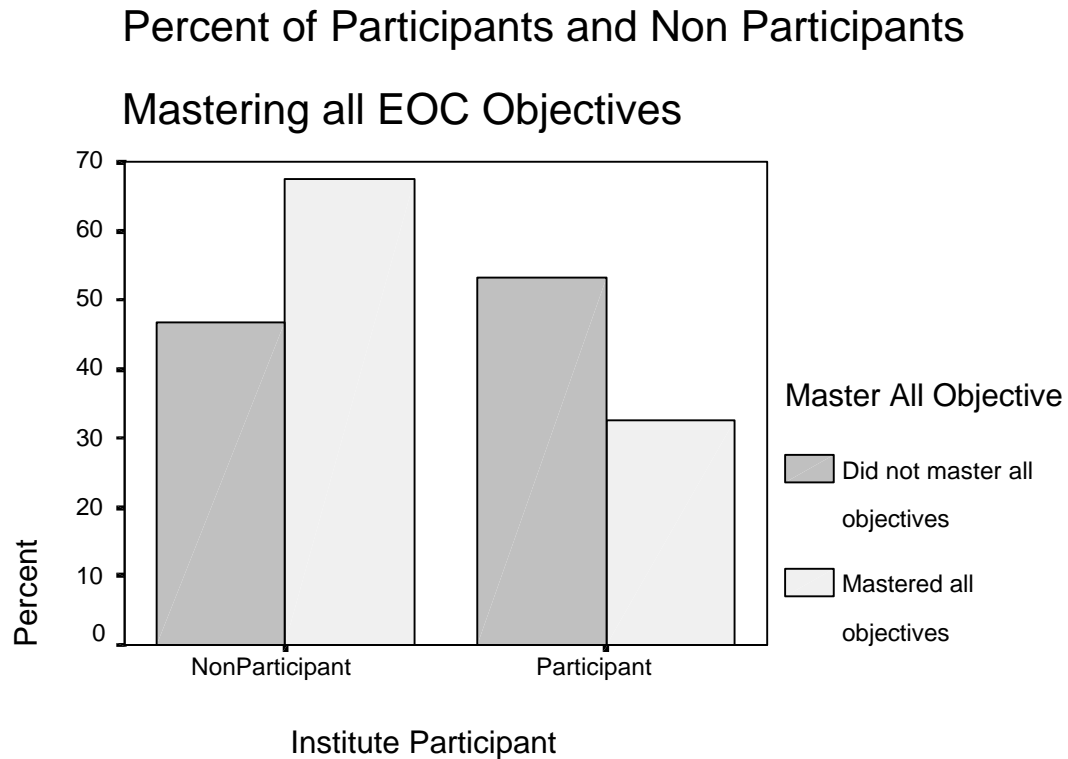
		Student Pass/Fail	
		Pass	Fail
Teacher	Institute Participant	Yes	No
	Non Participant	No	Yes

Mastering all EOC objectives was analyzed against teacher participation in the institute (see Figure 3). A statistically significant difference was found in the proportion of the scores for non-participants who mastered all EOC objectives

(29/64) and the scores of participants who mastered all EOC objectives (14/54).

The null hypothesis was rejected, with $\chi^2=4.75$, $N = 118$, $df = 1$, and $p < .05$.

Figure 3.



To further determine where the EOC frequencies differed, two additional chi square tests were analyzed with sorted groups. The null hypotheses for these tests stated that the differences across groups were the same across categories. When comparing the frequencies for participants and non-participants for mastery of all EOC objectives, the frequency for non-mastery was significantly higher for participants (40/54) than non-participants with $\chi^2 (1, N = 54) = 12.52$, $p < .001$. Although the frequency for non-mastery for non-participants was also high (35/64), it was not statistically significant. Additionally, when comparing the

frequencies of students who did or did not master all objectives to their teachers' participation in the institute, the non-participants' frequency for mastering all objectives (29/43) was significantly higher than for participants (14/43) with $\chi^2 (1, N = 43) = 5.23, p < .05$. In both tests, the null hypothesis was rejected. It would seem that the results of this test are contrary to the expected results. An examination of the class attributes of the non-participant teachers reveals a number of confounding variables. These include students who are not identified as being of low SES status, with gifted and talented abilities, and being at-risk for failure and for dropping out of school prior to graduation. These confounding variables are discussed in depth, later in the study.

The next chi square test was used to determine differences in the proportion for student mastery of all EOC objectives for students identified as GT or not identified as GT in non-participants' classes. There was a statistically significant difference in the odds ratio for GT students who did master all objectives (23/34) and those who did not master all objectives (11/34), with $\chi^2 (1, N = 64) = .03, p < .05$. The null could not be rejected in this test due to the small cell size ($N = 3$) for identified non-GT students in non-participants' classes. This situation occurred because based on district enrollment policies, students cannot enroll in PAP biology courses if they are not identified as gifted and talented. This variable was appropriately investigated using logistic regression later in the study.

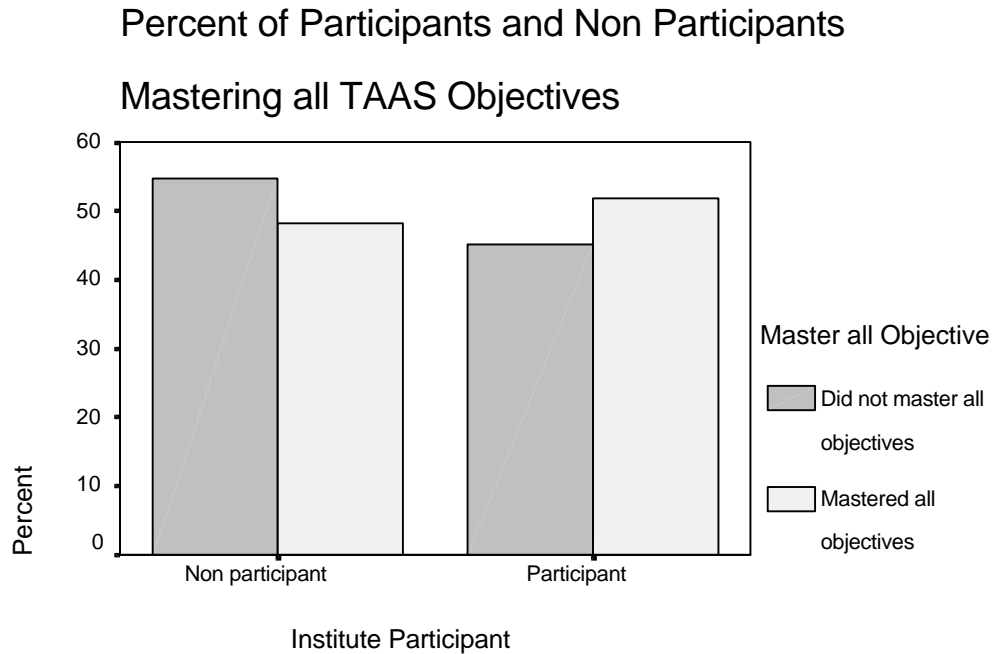
Because classes of convenience were used in the study, many non-participants' classrooms had a greater number of students who did qualify but did

not enroll in the GT programs. That in itself is not the issue. Of these non-GT students in the non-participants classes, 18/30 (60%) mastered all objectives compared to 12/30 (40%) who did not. For a variety of reasons, this student group included many who were not identified as gifted and talented, low SES, at risk, or needing student support services. By not being able to control for some potentially confounding variables for these student groups, it was not possible to determine the reasons for a higher mastery rate. Teachers in these groups use traditional instructional strategies addressed in the traditional professional development sessions that were aligned to meeting the needs for student achievement on the EOC examination. Instruction in classes identified for GT students typically do not emphasize skills needed for meeting minimum expectations, but spend more time addressing the higher-order thinking skills addressed on PAP and AP exams and in the Inquiry Institute.

To further investigate the mastery scores for the TAAS exam, a chi square test using cross tabulation was run to determine if a relationship existed between participation in the institute and mastering all TAAS objectives (see Figure 4). No other factors were controlled. The null hypotheses stated that the differences across the groups were the same. The results were not statistically significant, and the null hypothesis was not rejected. The chi-square test for equal proportions found there was not a statistically significant difference between the proportion of participants who mastered all TAAS objectives (29/43) and the proportion of non-participants who mastered all TAAS objectives (27/44) with χ^2

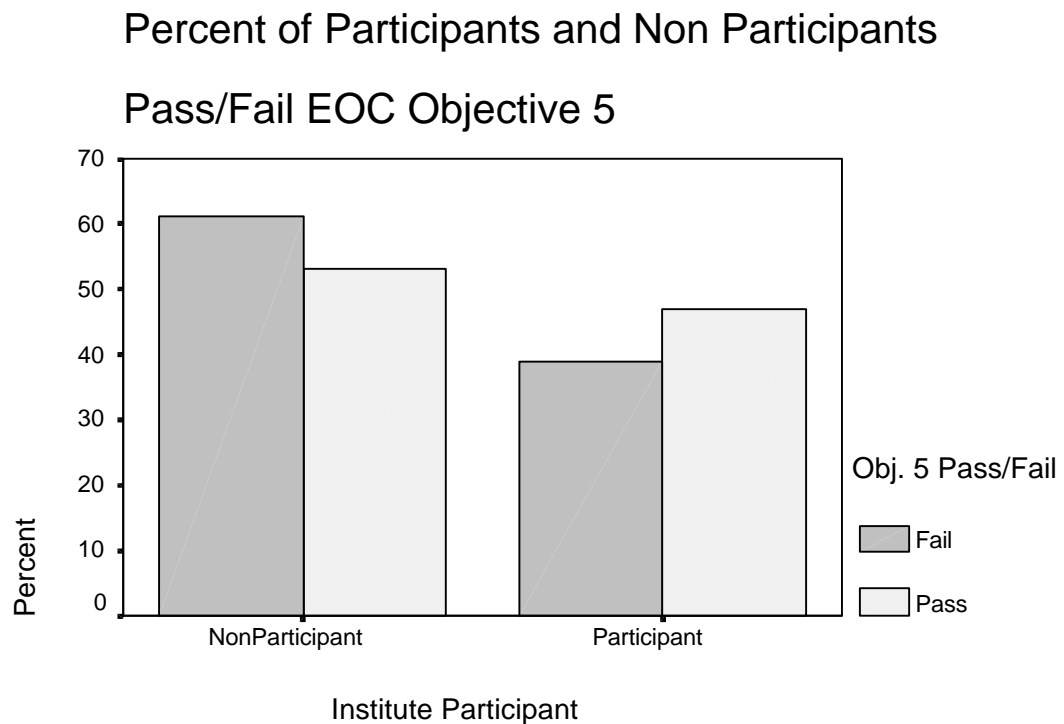
= .350, $N = 87$, $df = 1$, $p = .56$. Since there were no significant differences in f , no further tests were analyzed for mastery of all TAAS objectives.

Figure 4.



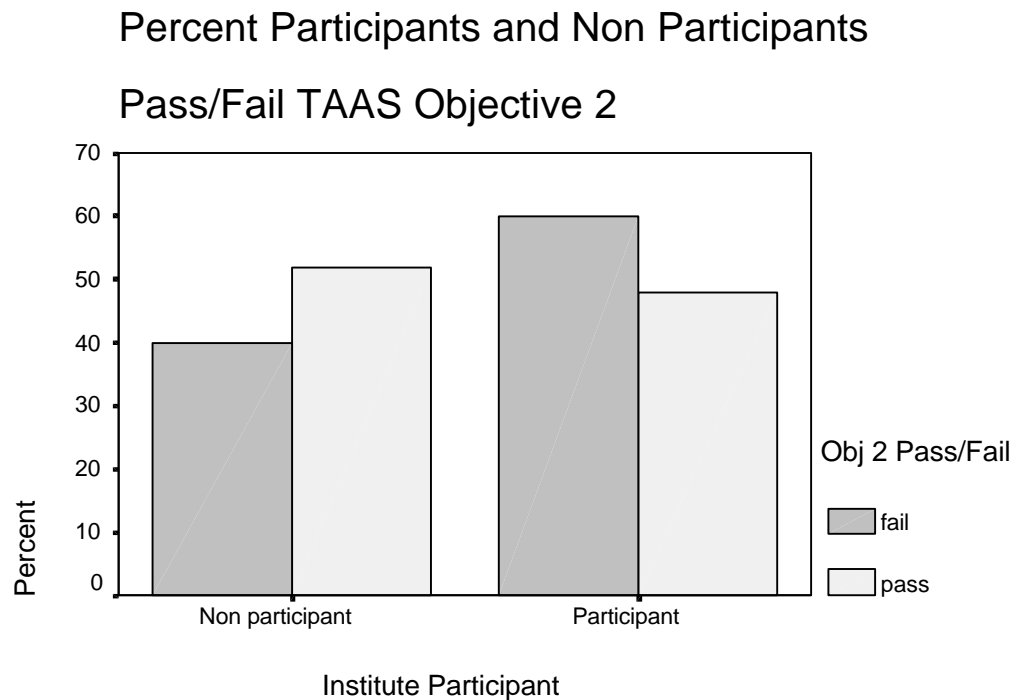
EOC objective 5 was designed to assess students' critical thinking skills. To determine the difference between the scores of student groups for teacher participation and non-participation in the institute and passing EOC objective 5, a second chi square test using cross-tabulation was run using EOC data (see Figure 5). The null hypothesis was not rejected for the chi square test. The proportion of the students for non-participants who passed EOC objective 5 (53/64) was not significantly different from the proportion of students for participants who passed EOC objective 5 (47/54), with $\chi^2 = .404$, $N = 118$, $df = 1$, and the $p = .53$. No further chi square tests were run because the results were not statistically significant.

Figure 5.



Finally, a chi square test using cross-tabulation was run to determine the overall difference in outcome based on a comparison of the proportion of students who passed the TAAS objective 2 and teacher participation. The proportion of participants' students who did pass objective 2 on the TAAS (37/43) was not significantly different from the proportion of non-participants' students (40/44) who passed TAAS objective 2 (see Figure 6). The null could not be rejected because of sparse cells. One cell had an expected count of less than 5; this exceeded the guidelines of no more than 20% of the cells can be smaller than 5. The chi square value was as follows: $\chi^2 = .505$, $N = 87$, $df = 1$, and the $p = .48$.

Figure 6.



Odds Ratio

Chi tests resulting in Pearson χ^2 was also used to predict the odds ratio for participants and non-participants mastering all objectives for TAAS and EOC, as well as passing EOC objective 5 (see Table 16). The null hypotheses stated that there would be no differences between the observed and expected frequencies caused by the independent variable (teacher participation or non-participation in the inquiry staff development) for passing specific objectives and mastering all objectives. The analysis was not conducted for passing TAAS objective 2 because of small cell numbers. For that cell, the validity of chi-square test would be suspect.

Table 16. Odds Ratio for EOC Mastering all Objectives/ Teacher Participation in Institute

		Mastered all objectives		Total
		Did not master all objectives	Mastered all objectives	
Non-participants	Count	35	29	64
	% within participation	54.7%	45.3%	100.0%
Participants	Count	40	14	54
	% within participation	74.1%	25.9%	100.0%
Total	Count	75	43	118
	% within participation	63.6%	36.4%	100.0%

The null hypothesis was rejected for the test comparing frequencies for mastering all EOC objectives and institute participation. The results revealed a Pearson χ^2 value of 4.75, $p = .03$, which was statistically significant at $p < .05$. Oddly enough, based on the odds ratio analysis, a student of a non-participant would have a better chance of mastering all objectives on the EOC (29/64) than a student whose teacher did participate in the institute (14/54). This could possibly be the result of the ceiling effect or confounding variables, as mentioned earlier, but it is worthy of further analysis. According to Huck (2000), confidence intervals that do not cross 1 would be significant to the .05 level, and the H_0 would be rejected. Based on values for the participants and non-participants mastering or not mastering all EOC objectives, the null was rejected (see Table 17). The value of 1 was not included in the 95% confidence interval. No other odds ratios for the EOC or TAAS binary dependent variables were statistically significant in the chi square tests.

Table 17. Confidence Intervals for Odds Ratio for EOC Participants and Non-Participants Mastering All Objectives.

	Value	95% Confidence interval	
		Lower	Upper
Odds ratio for did not master all /mastered all	.42	.19	.92
Non-participants	.74	.56	.97
Participants	1.75	1.03	2.96

Note. EOC total valid cases N = 118.

Logistic Regression

Cross-tabulation can be used to provide insights on questions such as “what happened” but not “why it happened” (Gaither, 1992). Multivariate analysis using logistic regression enables one to control for other variables. It is used to estimate how various factors influence the probability of the outcome. The purpose of this study is to examine the influence of student characteristics on their ability to pass objectives dealing with critical thinking, problem solving, and lab procedures, as well as their ability to show mastery of all objectives. The student categorical independent variables of gender, gifted and talented identification (GT), socioeconomic status (SES), at-risk identification, instruction by an institute participant VS. non-participant, and the continuous variable for ITBS scores were analyzed in two steps in this logistic regression study.

Once the model was set up in step 0, all independent variables were simultaneously added to the model, using the enter method in step 1. Categorical variables were coded 0 if the attributes were absent and 1 if the attributes or qualities were present. Males were coded 1 and females were coded 0. Sample

size of 118 for the EOC was adequate, based on the “Rules of Thumb” as established by Thorndike and cited in Peng, So, Stage, and St. John (2002). These rules include having at least 50 participants, plus 10 times the number of variables (100), and the sample size should be equal to 50 plus the square of the number of variables (100). The null hypothesis for the logistic regression states that all $\exp(\beta)$'s or odds ratios for the independent variables of the samples are equal to those of the population and are equal to 1. The Wald test was used to decide if there was any difference in the odds ratio of the dependent variable (mastering objectives or not mastering objectives) when affected by the added independent variables. The Wald test is tied to the theoretical distribution of χ^2 or the chi-square distribution (Huck, 2000).

The overall percentage for the logistic regression model based on mastering all EOC objectives (64.8) indicated more than a 50:50 probability in step 0, without the predictor variables entered (Wald $\chi^2 = 7.45$, $df = 1$, $p < .01$). The predictor variables were placed in the model simultaneously using the enter method (see Table 18). With the exception of the ITBS, none of the predictor variables were significant (Wald $\chi^2 = 12.51$, $p < .001$). The null hypothesis was rejected because the $\exp(\beta)$ was not equal to 1 in the population. The model detected a statistically significant difference ($p < .001$) for the students mastering all the EOC objectives and the scores on the ITBS exam. The odds ratio or $\exp(\beta)$ for ITBS indicated that students who did well on the ITBS exam in 7th grade were .05 times more likely to have mastered all EOC objectives than students who did not.

Table 18. Logistic Regression for Mastering All EOC Objectives

Variables	β	<i>df</i>	Wald χ^2	<i>p</i>	Exp (β)
Participation	.51	1	.79	.37	1.66
Gender	.74	1	1.43	.23	2.09
GT	.70	1	.71	.40	2.02
At Risk	.08	1	.00	.95	1.08
ITBS	-.05	1	12.51	.00*	.95
Constant	-4.20	1	.04	.85	.02

* $p < .001$

The logistic regression model was also used to determine the association of the five categorical and one continuous predictor variable for EOC objective 5 (see Table 19). The null hypothesis states that all exp (β)'s or odds ratios for the independent variables of the samples are equal to those of the population and are equal to 1. The overall percentage for the model with no variables entered (84.1), indicated more than a 50:50 probability. The step 0 model was also found to be statistically significant (Wald $\chi^2 = 32.64$, $p < .001$). The categorical independent variables were placed in the model simultaneously using the enter method. As a result of the logistic regression, the null hypothesis was rejected because the odds ratios did not equal 1. The ITBS performance was found to be statistically significant for predicting the probability to pass/fail EOC objective 5 (Wald $\chi^2 = 14.57$, $p < .001$). The odds for students who did well on the ITBS were .07 times more likely to pass EOC objective 5 than the odds for students who did not do well on the ITBS. These findings are not surprising. The criterion-referenced EOC exams assess science knowledge and skills at lower cognitive levels in 9th grade biology compared to higher levels of assessment for critical

thinking skills on the ITBS given in 7th grade. It would be expected that students who did well on the ITBS would continue to do well on the 9th Grade EOC.

Table 19. Logistic Regression for Pass/Fail EOC Objective 5

Variables	β	<i>df</i>	Wald χ^2	<i>p</i>	Exp (β)
Participation	.79	1	.71	.40	2.20
Gender	-.86	1	1.03	.31	.42
GT	-.82	1	.07	.82	.73
At Risk	.13	1	.01	.92	1.13
ITBS	.07	1	14.57	.00*	1.07
Constant	4.31	1		.91	74.26

* $p < .05$

Therefore, given the results of the logistic regression using the five independent variables, it is possible to suggest that there is an association for mastering of EOC objectives with scores on the ITBS, based on the correlation found in this analysis.

The second study used logistic regression to investigate student mastery of all TAAS objectives and the effects of the five categorical independent variables of participation, gender, GT, SES, at-risk, and the continuous independent variable for the ITBS scores. Because of the small sample size for low SES ($n = 3$), the results for this variable were not used. It could possibly cause the results of this study to be questionable, based on Thorndike's Rules of Thumb (Peng, et al. 2002). Long (1997) suggests that at least 10 observations per parameter used would seem reasonable for maximum likelihood (ML) models. Caution should be taken in assigning strength to the results from a small category in the logistic regression when there are less than 10 scores for a

parameter. The null hypothesis for the logistic regression for mastering all TAAS objectives stated that all β 's would equal one.

The logistic regression model for mastering all TAAS objectives had an overall percentage (67.1) that indicated a more than a 50:50 probability without any predictive variables added. The model also indicated statistical significance with Wald $\chi^2 = 8.53$, $df = 1$, $p = .001$. The null hypothesis was rejected when the variables were simultaneously added to the model in step 1 (see Table 20). The odds for students who were gifted and talented to master all TAAS objectives was 6.74 times more likely than the odds for students who were not GT. There was a statistically significant difference ($p = .005$) between the scores of GT and non-GT qualified students in mastering all TAAS objectives. Based on the model, it is possible to predict student mastery of TAAS objectives based on the GT qualifications of the student.

Table 20. Logistic Regression for Mastering All TAAS Objectives

Variables	β	df	p	Exp (β)
Participation	.15	1	.81	1.16
Gender	.10	1	.87	1.10
GT	1.91	1	.005*	6.74
At Risk	.94	1	.34	2.57
ITBS	.01	1	.30	1.01
Constant	-2.19	1	.20	.11

Note. SES cell contains less than 5 values for qualifying for services. At Risk contains less than 10. * $p < .01$,

The logistic regression model was also used to investigate the effects of the same predictor variables on passing TAAS objective 2 (see Table 21). The null hypothesis stated that there would be no difference between the observed and exp (β) for the entered dependent variables. The overall percentage (88.2)

for the model without variables entered indicated more than a 50:50 probability as a predictor model. The variables were entered simultaneously into the model, and the results for GT were again statistically significant ($p = .02$) as a predictor for TAAS objective 2. The null was rejected because $\exp(\beta)$ showed a difference in the samples. The odds for students who were gifted and talented were 13.91 times more likely to pass objective 2 than the students who were not. Because TAAS objective 2 focuses on critical thinking, inquiry, and problem solving, students who receive instruction in the use of inquiry, problem solving strategies, and the use of critical thinking skills would be more likely to be successful on this objective. Students in the gifted and talented program may have more opportunities to use and practice these higher-level skills compared to students receiving instruction in a regular classroom. When instruction centers on mastery of basic knowledge and skills, as in non-differentiated classrooms, students may not do as well on TAAS objective 2.

Table 21. Logistic Regression for Pass/Fail TAAS Objective 2

Variables	β	df	p	$\exp(\beta)$
Participation	-1.61	1	.10	.20
Gender	-.12	1	.89	.89
GT	2.70	1	.02*	14.91
SES	-8.18	1	.79	.00
At Risk	2.20	1	.07	9.04
ITBS	-.04	1	.14	.96
Constant	10.75	1	.73	46381.59

Note. SES cell contains less than 5 values for qualifying for services. At Risk contains less than 10. * $p < .05$.

In summary, the logistic regression model was used in the analysis of the six independent variables as predictors when placed in the model for the TAAS and EOC objectives. Table 22 is a summary of the likelihood ratio (LR) analyses of the factors that were statistically significant as predictors. When the student attribute for identified gifted and talented (GT) was added to the models for mastering EOC objectives, the p values were statistically significant ($p = .001$). When the GT predictor variables were added to the LR models for mastering TAAS objectives, they were also statistically significant ($p < .01$ and $p < .05$). It is interesting to note that, based on the results of this analysis, it can be accurately predicted that students who are gifted and talented will do well on the ITBS science subtest in 7th grade and will perform well on TAAS and Biology EOC exams. Exams such as the TAAS and EOC may not be effective in detecting statistically significant differences or practically significant effects in the population selected for this study. Differences in test scores that are associated with teacher attributes, student factors, or the teachers' professional development could not be adequately analyzed due to small cell sizes in some subpopulations that research has indicated are most responsive to the strategies being investigated.

Summary

The results of the study showed a stronger relationship between student achievement on standardized exams and (1) prior knowledge and (2) being identified as gifted and talented rather than with teacher participation in an

Table 22. Summary of Statistically Significant Independent Variables in Logistic Regression Analyses of TAAS and EOC Data.

Dependent variables	Independent variable	β	df	p	Exp (β)
EOC mastered all objectives	ITBS	-.05	1	.00**	.95
EOC pass/fail objective 5	ITBS	.07	1	.00**	1.07
TAAS mastered all objectives	GT	1.91	1	.005**	6.74
TAAS pass/fail objective 2	GT	- 2.70	1	.02*	14.91

* $p < .05$. ** $p < .01$.

inquiry-based professional development program. When student achievement was analyzed with student's prior abilities and knowledge held constant, it would seem logical to assume that gains in achievement may be attributed to differences in instructional practices. But, these assumptions did not hold true in the study. Based on the results of the study, prediction of student achievement on the TAAS may be possible based on students' gifted and talented qualifications. ITBS scores may possibly be the best predictor of student achievement on the EOC.

When determining whether a relationship exists between student achievement and student attributes such as gender, low SES status, or being classified as at-risk, again teacher participation in an inquiry-based professional development could not be given credit for the gains made. Instructional strategies proven to be effective in the instruction of these populations could not be evaluated, and the effect of institute training could not be measured due to small sample size. Test scores for objectives, which typically assess effects of higher-

level instruction on subpopulations, did not show gender effect. Small sample sizes did not allow for comparisons among the other groups.

This study did reveal, however, the effects of confounding variables, which cloud the accuracy of the results. The instructional needs of students in regular classes are often unidentified. Students are identified as GT or needing Special Education based partially on their ITBS scores. Students with scores falling between both ends of the ITBS scale have a large range of needs that must be addressed in the regular classroom. The most efficient means of meeting these needs have been the foundation of traditional instruction emphasized in traditional staff development programs. For those students who qualify but elect to not be identified or served by GT or Special Education classes, the inquiry instructional strategies taught through the institute could have contributed to their success. But, in this study, the effect could not be analyzed properly.

Finally, the ceiling effect had a confounding impact on the conclusions drawn from the studies. When such a large number of students pass an achievement test meeting both the minimum expectations and high levels of mastery, there is reason to believe that the test is not measuring all the levels of learning or instruction taking place in the science classroom.

Conclusion

The data analysis for this study included collection and coding of the data for each analysis method. Once the data were coded, analysis using various statistical methods produced statistically significant results in some instances, which led to the rejection of several of the null hypotheses. Both the statistical

and practical significance of the findings were reported. Actual sample sizes shrank from the time of the proposal due to missing or incomplete data.

Statistical analysis based on small sample data sets were carefully studied and subjected to further analysis prior to rejection or retention of the null to avoid a Type I error. This chapter outlined the systematic study of the data collected and its subsequent analysis.

In chapter 5, the analyzed data are used to make generalizations from the sample to the population based on the assumptions for each method. The statistical significance of the variables and the practical significance based on effect size are reviewed. The discussion of the findings is used to study future professional development and instructional programs that will have an impact on student achievement. Further generalizations are made about the predictive value of student attributes and scores of previously administered standardized assessments. The findings may potentially impact the use of standardized exam scores in shaping professional development for instructional practices that attempt to “raise the bar” on the expected levels of achievement. The discussion section will also address the future of professional development on the design and implementation of curriculum, instruction, and assessment in science public education.

CHAPTER 5

DISCUSSION

Introduction

This chapter consists of a discussion of the results of the study. Highlights of each portion of the study include a brief summary of the data, which supports the hypotheses, determination of achievement of the purposes, as well as references to the literature review, which show the consistency of this study within that body of research and theory. An interpretation of the findings, as well as a discussion of the possible implications, is based on the strengths and limitations, of the study. The chapter closes with recommendations for further research in this area.

Limitations and Generalizability of the Results

This study included students from two high schools and two middle schools from a suburban school district in a north Texas metropolitan area. The biology teachers were from the two different campuses. The 8th grade teachers were paired from the same campuses. The control and experimental groups were selected from intact 8th grade science classes and 9th and 10th grade biology classes. The quasi-experimental design for this study was limited to the students of the three biology teachers and the two eighth grade science teachers who participated in the Inquiry Institute. Control and experimental teachers were matched by years of teaching experience, level of instruction, area of

certification, and degrees earned. Students taking the 8th grade science TAAS were in the eighth grade and not in Special Education. Students taking the biology EOC were enrolled in the 9th or 10th grade biology and were not in Special Education. All biology students were ninth graders with a few tenth graders. Students without matching ITBS scores were eliminated from the study. Special Education students were not included in the data collection reported by the TEA reports. Special Education status was based on the criteria identified in the Special Education Admission Review and Dismissal Committee's recommendations, which stated that the test was not the appropriate assessment for the level of instruction that the students were receiving.

8th grade science TAAS Discussion

The purpose of the TAAS study was to determine if teacher participation in an inquiry-based science teacher professional development program would impact student achievement on the 8th grade science TAAS test given in the 2001-2002 school year. The research hypothesis stated (H_1) that the 8th grade science TAAS scores for students of Inquiry Institute teachers would have higher mean scale scores than students of non-Inquiry Institute teachers. Upon first examination of the raw data, institute participants' students academic achievement appeared to be in agreement with the research or directional hypothesis. But, in order to view the actual effects of instruction for the 2001-2002 school year, prior knowledge in science was negated through the use of 7th grade ITBS science subtest scores. The mean scale scores for the 38 students (in PAP and regular science) instructed by teachers who participated in the

Inquiry Institute did not show a significant difference over the mean scale scores of the 38 students in classes taught by teachers who did not participate in the institute. Students who were identified as gifted and talented using ITBS scores as one indicator had higher mean scores than the general student population. This was true without regard to the professional development training of their teachers.

It is interesting to note that 97% of all regular 8th grade science students in the school district achieved minimum expectations on the TAAS, compared to the state average of 93%. These results could be an indication of the overall high level of instruction used in 8th grade. These results could also be an indication of the high caliber of students who are enrolled in the district, and who bring with them to 8th grade a rich background of science knowledge and skills as measured by the ITBS scores. Results could also be an indication of the ceiling effect addressed by Gall et al. (1996) who stated “when the range of difficulty of the test items is limited and therefore scores at the higher end of the possible score continuum are artificially restricted” (p. 533). As stated in the guidelines from TEA, “The goal of the assessment program in Texas is to measure student progress towards achieving academic excellence. The primary purpose is to provide accurate measurement in the areas of reading, writing, mathematics, social studies, and science. The results are used as a gauge for institutional accountability” (TEA, 1999, p.1). To that extent, the 8th grade science TAAS does accomplish that ends in assessing minimum expectations. It does not assess, however, the higher levels of critical thinking and problem solving required by the

TEKS and the National Science Benchmarks and modeled through the Inquiry Institute professional development program.

A review of the analysis of TAAS scale scores also indicated students enter 8th grade science with a strong background of science knowledge and skills based on the mean score on the ITBS of 71% and its high correlation to the TAAS test. It was interesting to note that there was a statistically significant difference in the mean scale score for GT students taught by participants compared to GT students taught by not-participants (1798:1748). This difference could be attributed to the more rigorous inquiry-based instructional strategies used in the PAP 8th grade classroom and modeled through the Inquiry Institute program. The PAP 8th grade science curriculum follows the TEA guidelines for secondary science courses (i.e. 40% labs and field experiences). These experiences encourage inquiry-based activities, critical thinking skills, problem solving, and authentic assessments. Inquiry Institute participants, through the process of immersion, had the opportunity to experience inquiry investigations, plan the implementation of the inquiry labs into their own classrooms, as well as network with other participants as students experienced the inquiry process through the labs and field experiences. It would be interesting to consider the types of teachers who respond positively to workshops structured in a pattern similar to the Inquiry Institute. Are teachers of GT students more motivated or more capable of applying the techniques discussed and modeled in the institute? Because of the limitations set by the knowledge level of the test items, the scale

scores could not differentiate the higher levels of critical thinking and instruction that took place in GT classrooms.

The purpose of the second study, using the results for the TAAS exam, was to take a closer look at the level of achievement on test objective 2 which assessed students' abilities in critical thinking and problem solving. It was hypothesized (H₂) that the 8th grade science TAAS scores for students of Inquiry Institute teachers would have higher passing rates on objective 2 than students of non-Inquiry Institute teachers. This study showed an 86% overall mastery of objective 2 regardless of their teachers' participation in the institute. Student gifted and talented identification did have a strong relationship with mastery of objective 2 items, again without statistically significant differences between participants and non-participants' classes. Because of the high passing rate and the small number of students who did not pass objective 2 (12%), it was impossible to judge if teacher participation in the institute had any significant effect on the regular student in the 8th grade.

Although the 8th grade science program in this study was not kit-based, the Inquiry Institute for professional development in this district modeled the use of inquiry-based instructional strategies, high quality curriculum, and multiple resources. The four-year study by Klentschy et al. (1999), found that students using inquiry-based lab strategies through a hands-on, kit-based program had higher achievement test scores. The study reported a 14-percentile point difference on the Stanford Achievement Test, 9th Edition, form T.

According to Sparks (1996), the quality of any professional development should be reflected in increased student achievement. Loucks-Horsley et al. (1998) indicated that the immersion approach to inquiry-processes in professional development is effective in improving student achievement. According to St. John (1999), professional development should also be reflective of the kind of instruction called for in the standard-based reform efforts. The Inquiry Institute addressed these research findings through the process and context of inquiry-based science.

Because of the ceiling effect of the 2002 8th grade science TAAS, it was impossible to determine the effects of instruction that go beyond the minimum standards set by the TEKS. Although not all students mastered objective 2, the high percentage of those who did show mastery would indicate that teachers are teaching to the level that is expected by the test items. It would be interesting to repeat this study with a more diverse population, balanced between students who qualified for GT programs and those who did not, to see the effects of inquiry instruction. It would be of value to repeat this study using the same population to determine if the new TAKS test is able to effectively assess knowledge and skills at the higher levels called for by reform efforts and the TEKS. Finally, based on the timeline for the Inquiry Institute, teachers were not able to fully implement the inquiry-based strategies prior to administration of the 8th grade science TAAS in 2002. It would be of value to repeat the study in succeeding years, using the same 8th grade science TAAS test, to determine the

effects of instruction by institute teachers who fully implement the strategies, prior to testing.

Biology End-Of-Course Exam Discussion

The purpose of the first part of the EOC study was to determine if teacher participation in an inquiry-based science teacher professional development program would impact student achievement on the biology EOC exam given to 9th and 10th grade biology students at the end of the 2002 school year. It was hypothesized that (H₃) EOC exam scores for students of Inquiry Institute teachers would have higher mean scale scores than students of non-Inquiry Institute teachers. Of the 118 students originally included in the study, 30 students did not have matching ITBS test scores to be used for equating the scores. The 25% rate of missing ITBS scores is most likely a reflection of the rate that students transfer into and out of the district, which is not unusual for the metropolitan area in which this district is located.

As with the results of the non-equated TAAS test, the results of the EOC analysis showed statistically significant higher mean scale scores for students of participants than for students of non-participants. But, when equated using the mean ITBS scores, the difference was not statistically significant between the two groups. The ITBS test was positively correlated with the EOC and revealed a large effect size towards the EOC scale score. This would indicate that students who do well on the ITBS in 7th grade would have prior knowledge that is being assessed on the EOC. It would also indicate that more than half the knowledge and skills assessed on the EOC was gained prior to entering 7th grade.

Prior knowledge and experiences, as studied by brain-based learning theorists (Jensen 1996) as well as cognitive and constructivists learning theorists (Caine & Caine, 1994), are the basis for learning and retaining new information. As new learning is linked to prior knowledge, the retrieval of old and new information is facilitated. The science background gained during the elementary years is measured on the ITBS 7th grade science subtest. Science experiences that are rich in hands-on, student-centered inquiry, whether gained in the school setting or through a child's own experiences will provide that needed prior knowledge and be indicated as higher ITBS test scores. Science experiences at the secondary level are more abstract and require higher levels of critical thinking skills. According to the theories of Vygotsky, Piaget, and Bruner, unless younger students are provided experiences at a concrete level on which to build a scientifically literate knowledge base, they may be unable to successfully manage learning at a more abstract level. Lower test scores on the secondary science exams that assess higher levels of critical thinking, may possibly be an indication of a lack of science experience at the elementary level, rather than a lack of understanding of scientific content knowledge taught at the secondary level. The EOC results may indicate an identified need for more science experiences at the elementary level, for the improvement of secondary science test scores.

The mean scale score for the EOC within the study was higher for participants than for non-participants (1879:1808). The overall district-mean scale score was higher than for the state, at 1759 and 1664 respectively. The state

standard of 1500 is required for meeting minimum expectations. In the district studied, of all students taking the EOC, 92% passed the exam, meeting minimum expectations based on the scale score. At the state level, only 80% of the students taking the EOC gained minimum expectations. It would again appear that the ceiling effect should be considered in the implications of these results. The level of the EOC may be limited and unable to adequately assess the higher levels of knowledge and skills that (a) the students are able to achieve and (b) the TEKS and National Science Benchmark standards require. The results of this study are not adequate to assess the effects of traditional or inquiry-based staff development based on differences in the EOC scale scores in student groups.

When reviewing the mean scale scores for students who were GT qualified in the classes of institute participants, it was interesting to note that it did not reveal the same effect size as was noticed in the 8th grade TAAS study. Students who were not identified as GT, but who were in participants' classes, had higher mean scores than those who were identified as GT but not in institute participants' classes. When students whose ITBS scores are high enough to qualify as GT, but are not identified as GT on teachers' rolls, confounding variables are created that may lead to variances not accounted for. Although the results could not be used due to small cell sizes, it would be of value to further investigate the effects of inquiry-based instructional strategies on the academic achievement scores of students, who are identified as GT but are enrolled in classes at the regular level rather than at the PAP level.

Further analysis of the EOC scale scores revealed no statistical difference in the adjusted scale scores for student demographic attributes such as low SES or identified at-risk status. The percentage of identified students in the population was small, as were the cell sizes in the study. Many students are identified as low SES status at the elementary level and the effects of academic interventions have been documented as higher levels of achievement on state mandated tests. At the secondary level, it is possible that the stigma attached to being identified in the low SES and at-risk categories may lead to under-identification and subsequent lack of intervention. These factors may add to the confounding variables that are apparent in this study.

Based on the research conducted by Coleman et al. in 1966 as described in Marzano et al. (2001), “the majority of the differences in student achievement can be attributed to factors like student’s natural ability or aptitude, socioeconomic status of the student, or the student’s home environment” (p.2). In a similar study by Jencks et al. (1972) it was found that differences in test scores were “due to factors the school cannot control” (p. 2). The results of this study disagreed with those findings. The results of this study agreed with the findings from Oakes et al. (1997), which stated that the differences in the achievement remain constant or improve when students are placed in heterogeneous groups. Across this district, all student populations met minimum expectations based on scale scores, and in this study there were no significant differences for student achievement scores based on SES or at-risk status.

According to Bracey (1997) it is reported that as students matriculate through the public school system in the United States, science achievement scores decline based on worldwide standards. The data provided by Greene (2002) about the TIMSS-R study, stated that the United States students' achievement is competitive in 4th grade science and mathematics. It begins to decline in 8th grade, and by 11th grade, the United States is no longer among the world leaders. In the United States, gender effects as well as the effects of SES become evident. It would be of value to study the effects of inquiry-based instruction on these test results to determine if the instructional strategies provide the difference required to close the gender gap. In this study of EOC scale scores from the perspective of gender effect, it was found there were no statistically significant differences in the EOC scale scores of males and females across teachers who used strategies from either traditional professional development or the Inquiry Institute. As indicated by the results of the TIMSS and TIMSS-R studies (Greene, 2002), gender effects in the United States are not a significant issue at the 8th grade level but are at the high school level. It would be of value to determine the gender effects for science achievement in the higher levels of science instruction. Studies conducted using the results of the newly developed 10th grade science TAKS and 11th grade Exit Level Science Exams may reveal a pattern that is not evident in the TIMSS and TIMSS-R data.

The purpose of the second part of the study using the EOC results was to take a closer look at the level of achievement on test objective 5 that assesses biology students' abilities in problem solving, critical thinking, and laboratory

procedures. It was hypothesized (H₄) that 9th and 10th grade biology End-Of-Course exam scores for students of Inquiry Institute teachers would have higher passing rates on objective 5 than students of non-Inquiry Institute teachers. The results of this study indicated an overall passing rate of 85%. Statewide, the passing rate was 44%. There was not a statistically significant difference between the passing rate for those whose teachers participated in the inquiry workshop and those who did not. Based on the standard set by the TEA for mastery of this objective, 5 out-of 6 items needed to be answered correctly. An analysis of EOC test questions revealed levels of skill and knowledge at the lower end of the critical thinking hierarchy. The student expectations for the TEKS are listed at higher levels than they are assessed on the current EOC. Based on the level of current instructional practices used by the teachers in this study, students are adequately prepared for the level of assessment presently used. According to the Dana Center (1998), the revised state assessments will ensure implementation of the TEKS and will change the focus of local curriculum and instruction. Currently, inquiry-based science programs are not needed to insure student achievement on the EOC. It could be of value to study the results of the revised 10th grade science TAKS test and the 11th grade EXIT science test to determine the level of critical thinking required to achieve objective 5 mastery.

Prediction of Student Mastery

On TAAS and EOC Objectives

The purpose of the last part of this study was to determine if teacher participation in the inquiry-based professional development program would be a

statistically significant means of predicting student achievement on the TAAS and EOC exams based on mastery of all objectives and specifically TAAS objective 2 and EOC objective 5. It was hypothesized (H_5) that students in identified subgroups, when taught by Inquiry Institute teachers, would have higher TAAS and EOC objective mastery rates than when taught by non Inquiry Institute teachers.

Through the use of odds ratios, it was possible to predict student mastery of TAAS and EOC objectives based on scores of the ITBS test and identified characteristics of the subgroups. Based on the results of the data for mastery of all EOC objectives and the data for each subgroup, the strongest predictor of student mastery was the ITBS score. Students who did well on the 7th grade exam were .05 more likely to have mastered all objectives than those who did not. Teacher participation in the institute, gender, GT qualification, and at-risk identification did not show any predictive values. The same student attributes were placed in the model for prediction of mastery of EOC objective 5. Again the strongest predictor of student mastery of all EOC objectives was the ITBS 7th grade exam. Students who did well on the ITBS were .07 times more likely to master all EOC objectives, than those who did not do well on the ITBS.

The model was then used to determine the predictive value for teacher participation in the institute on mastery of all TAAS objectives. Through the use of odds ratios, it was revealed that students who are identified as GT are 6.74 times more likely to master all TAAS objectives than those who are not. Neither teacher participation, gender, nor at-risk identification were significant predictors.

For prediction of mastery for TAAS objective 2, it was not surprising to find GT identification was the statistically significant predictor. Students who are gifted and talented are 13.91 times more likely to master TAAS objective 2 than students who are not GT qualified.

Due to the natural inclination of GT students, and the low student expectations of the TAAS and EOC exams, teachers are able to use a variety of traditional instructional strategies that are efficient but that do not challenge the critical thinking skills of GT students. The present assessments are not of sufficient rigor to adequately assess the levels of knowledge and critical thinking skills GT students are capable of attaining. It would be of value to conduct this research using student populations with a wider range of student attributes.

Based on informal observation, Inquiry Institute teachers may not have been fully implementing the inquiry technique during the instructional year. One of the original assumptions in the study was that teachers would immediately implement the instructional strategies gained while attending the Inquiry Institute. Based on the studies by Hord et al. (1998) and Loucks-Horsley et al. (1998) in the use of the Concerns Based Adoption Model (CBAM) for the implementation of change, time is needed for teachers to adjust to the changes inquiry-based instruction would entail. The CBAM research found that

- The change process is an extremely personal experience and how the individual perceives it will greatly affect the outcome.
- Individuals progress through various stages regarding their emotions and capabilities regarding the innovation. (Hord et al, 1998, pp. 5-6)

The last session of the institute that dealt specifically with the implementation of inquiry and how to move from traditional teaching methods to inquiry-based teaching methods did not occur until just a few days before the administration of the EOC and 8th grade science TAAS exams. Teachers did not have the opportunity to fully implement the training or experiences they gained from the institute. It would be expected, as evidenced by the research of Hord et al. (1998) and Loucks-Horsley et al. (1998), that achievement test scores of future students of teachers who participated in the institute could show gains based on the full implementation of inquiry-based instructional strategies. According to the CBAM model, if teachers were supported through a concerns-based professional development program and resource materials, institute participants would be more inclined to implement the change in their instructional practices.

In a review of instructional strategies that meet the needs of all students, research highlights the application of the best practices as advocated by Lezotte (2002). These practices are based on a collection of education research findings, that provide support to the argument that *how* content is taught is equally as important as *what* is taught. The effective schools literature reveals strategies that empower schools and districts to implement the long term, systemic changes needed to ensure that every student succeeds. It would be of value to repeat this research based on the data from the revised tests and implemented best practices for all science instruction.

The results of this study may be generalizable to most public school districts with similar populations, whose levels of student achievement are being

assessed by state-mandated achievement tests. Missing ITBS scores eliminated data for many students in low SES and at-risk status subgroups. As a result, these groups were not well represented in the sample. Generalized findings based on the data from this study must be used with caution and only applied to similar populations. Further research is needed to determine the extent to which these results generalize to students of various subpopulations (e.g. ethnic groups, low SES and at-risk status) who are instructed using Inquiry-based instructional strategies. Replication of the study would be important.

Summary

It was the purpose of this study to analyze the relationship between inquiry-based science professional development and student achievement on state-mandated science standardized tests. Based on the imperatives established by the National Staff Development Council, if the value of professional development is measured by gains in student achievement, then there should be a relationship between the content, process, and context of professional development programs and the scores on the TAAS and EOC. The purpose of professional development is to provide teachers with skills and resources to close the gaps between student learning objectives as determined by needs assessments and aligned assessment objectives. Based on the CBAM model (Hord et al., 1998), time is needed for teachers to fully implement the strategies gained during professional development into instructional practice. When student achievement improves, and when professional development is focused on closing this student achievement gap, Sparks (2002) contends there

is a direct relationship to the professional development program that is aligned. When the content and level of the assessment is not aligned with the content or level of the student learning objectives, then the effects of the professional development cannot be accurately measured. According to Sparks, the highest measure of effective professional development is gains in student achievement.

This study analyzed student scores from the TAAS and EOC exams with ITBS scores as covariate to determine the extent that an inquiry-based professional development program was able to make a difference in student achievement on test objectives that assessed the critical thinking skills and problem-solving strategies modeled in the institute. It was hypothesized that the students of teachers who participated in the process and strategies of inquiry would have higher scale scores on the TAAS and EOC exams and better rates of mastery on the TAAS and EOC objectives that assess higher critical thinking skills. It was also hypothesized that it would be possible to predict student achievement on the TAAS and EOC critical thinking objectives based on teacher participation in a science inquiry institute. The results of this study indicate that professional developers might be unable to determine the effects of inquiry-based professional development programs, if any, by analyzing the results of the standardized tests currently being used in Texas. Because of factors such as the ceiling effect and test items that are intended to but do not address critical thinking skills, the 8th grade science TAAS and biology EOC tests are limited to a level of science knowledge and skills that cannot accurately assess the implementation of the content, processes, and context of the Inquiry Institute.

This study revealed that time must be given for the full implementation of any professional development program prior to student assessment, if the relationship between professional development and student achievement are to be fairly tested. The results of this study also indicate the best predictor of student mastery of objectives on the 8th grade science TAAS and biology EOC is prior knowledge as demonstrated on the 7th grade ITBS science subtest.

When the standardized tests items are revised and adjusted to reflect higher-level student expectations, the effects of sophisticated instructional strategies and problem-solving applications might then be accurately assessed. A needs assessment based on the gaps between student achievement and curriculum standards could highlight the need for professional development. Through inquiry-based professional development, strategies that are based in inquiry and that teach to higher levels of critical thinking could be provided. It was in anticipation of these assessment revisions that the Inquiry Institute program was offered.

When using standardized tests to measure the impact of professional development, it would be important to guarantee that the testing instrument selected is accurately calibrated to measure content and levels of instruction not limited to the minimum standards. When state and national standards call for higher levels of student achievement, when a standards-based curriculum is taught, when professional development prepares teachers with the skills required for instruction of the standards and student achievement of the standards, and when time and support is given for full implementation of the program, the only

accurate assessment of the impact of professional development would be increases in or higher student achievement. That determination is not possible if the assessment instrument is not aligned to the content and level of the standards.

Conclusion

This chapter consisted of a discussion of the results of the study. Highlights of each portion of the study included a brief summary of the data analyses that supported the decision to either reject or retain the null hypotheses, determination of achievement of the purposes, as well as references to the literature review, which showed the consistency of this study within that body of research and theory. An interpretation of the findings, as well as a discussion of the possible implications was based on the strengths and limitations of the study. The chapter closed with a summary and recommendations for further research in this area.

REFERENCES

- AERA.net, (2000). *AERA position statement concerning high-stakes testing in prek-12 education*. Education Research Network. Available online from: www.aera.net/about/policy/stakes.htm.
- American Association for the Advancement of Science, (1989). *Science for all Americans*. Annapolis, MD: Project 2061. Available online from: <http://www.project2061.org/tools/sfaaol/sfaatoc.htm>
- American Association for the Advancement of Science, (1993) *Benchmarks for science literacy*. Annapolis, MD: Project 2061. Available online from: <http://www.project2061.org/tools/benchol/bolframe.htm>
- Behar-Horenstein, L.S. (1999). Narrative, teaching, and teacher thinking. In H.C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (229-248). Berkley, CA; McCutchan Publishing.
- Bracey, Gerald. (1996) The sixth Bracey report on the condition of public education, *Phi Delta Kappan*. October, 127-138.
- Bracey, G. (1997). *Setting the record straight: Responses to misconceptions about public education in the United States*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Brandt, Ronald S. (2000). Introduction: A new century. In S. Brandt (Ed.), *2000 ASCD yearbook: Education in a new era*, Alexandria, VA: Association for Supervision and Curriculum Development. Available online from: www.ascd.org/readingroom/brandt00.html

- Bruner, J., Goodnow, J. & Austin, G. A. (1967). *A study of thinking*. In G. Caine & R. N. Caine, (1991) *Making connections: Teaching and the human brain*. Menlo Park, CA: Innovative Learning Publications.
- Carnoy, M., Loeb, S. & Smith, T.L., (2001). *Do higher state test scores in Texas make for better high school outcomes?* University of PA: Consortium for Policy Research in Education. Available online from: www.cpre.org/publications/rr47.pdf.
- Charles A. Dana Center (1998). What it means to implement the TEKS: The role of local curriculum development. *Issue Brief*, 1(1).
- Charles A. Dana Center (2002a) *Instruction and assessment: Frequently asked questions*. Austin, TX: University of Texas at Austin. Available online from: www.tenet.edu/teks/science/stacks/instruct/faq.html
- Charles A. Dana Center (2002b) *Program development: Elements of an effective science program*. Austin, TX: University of Texas at Austin. Available online from: www.tennet.edu/teks/science/stacks/progdev/elements.html
- Chiappetta, E. L. (1997). Inquiry-based science: Strategies and techniques for encouraging inquiry in the classroom. *The science teacher*, 64(11), 22-26.
- Commission for Excellence in Education (1983). *A nation at risk*. Washington, D.C: National Commission for Excellence in Education. Available online from: www.ed.gov/pubs/NatAtRisk/risk/html
- Darling-Hammond, L. & Snyder, J. (1992). *Curriculum studies and the traditions of inquiry: The scientific tradition*. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 41-78). New York: Macmillan Publishing.

- Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappa*. Available online from:
www.pdkintl.org/kappan/darling.htm
- Darling-Hammond, L. (1999). Professional development for teachers: setting the stage for learning from teaching. *Teaching and California's future*. Santa Cruz: The Center for the Future of Teaching and Learning. Available online from: www.cft.org/documents/Jolly-paper.pdf
- Department of Education (1998). Goals 2000: History the legislation. *Goals 2000: Reforming education to improve student achievement – April 30, 1998*. Available online from: www.ed.gov/pubs/G2Kreforming/g2ch1.html
- Eisner, E. W. (1999). The uses and limits of performance assessment. *Kappan*, Available online from: www.pdkintl.org/kappan/keis9905.htm
- English, F. (2000). *Deciding what to teach and test: Developing, aligning, and auditing the curriculum*. Thousand Oaks, CA: Corwin Press.
- Evertson, C.M. & Randolph, C. H. (1999). *Perspectives on classroom management in learning-centered classrooms*. In H.C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (249-268). Berkley, CA; McCutchan Publishing.
- Exploratorium, (2002). *Workshops and programs*. San Francisco, CA. Available online from: <http://www.exploratorium.edu/>
- Fraser, B. J. (1999). "Grain sizes" in learning environmental research: Combining qualitative and quantitative methods. In H.C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (285-296).

Berkley, CA; McCutchan Publishing.

Fullan, M. G. (1994). *Coordinating top-down and bottom-up strategies for educational reform*. In R. F. Elmore & S. H. Furhman (Eds.), *1994 ASCD yearbook*. Association for Supervision and Curriculum Development.

Available from: www.ascd.org/readingroom/books/elmore94toc.html

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*, (6th ed.). White Plains, NY: Longman Publishers, USA.

Gaither, G. H. (1992). Persistence patterns in public higher education: The case of Texas. *College and University* (67)1, 245-52.

Glickman, C. D. (2001). Holding sacred ground: The impact of standardization, *Educational Leadership*. Alexandria, VA: Association for Supervision and Curriculum Development. Dec.-Jan, p. 46-51.

Goldberg, M. and Harvey, J. (1983). A nation at risk: The report of the national commission on excellence in education, *Phi Delta Kappan*, Sept, 14-18.

Greene, B. D., Herman, M. & Haury D. L., (2002). *TIMSS: What have we learned about math and science teaching?*, ERIC Clearinghouse for Science, Mathematics, and Environmental Education. Available online from:

www.ericse.org/digests/dse00-05.html

Guskey, T.R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.

Guskey, T. R. & Sparks, D. (1996). Exploring the relationship between staff development and improvements in student learning. *Journal of Staff Development*, 17(4). Available online from:

www.nsdcd.org/library/jsd/f_gusky.html

- Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences*, (4th ed.). New York: Houghton Mifflin Company.
- Hord, S. M., Rutherford, W.L., Huling-Austin, L., & Hall, G.E., (1999). *Taking charge of change*. Austin, TX: Southwest Educational Development Laboratory.
- Huck, S. W. (2000). *Reading statistics and research*, (3^d ed.). Menlo Park, CA: Addison Wesley Longman, Inc.
- Jensen, E. (1996). *Brain-based learning*. Del Mar, CA: Turning Point Publishing.
- Joyce, B. & Weil, M. (with Calhoun, E.) (2000). *Models of teaching* (6th edition). Boston: Allyn and Bacon.
- Kaufman, R. A. (1983). *Needs assessment*. In F. English & O.L. Davis (Eds.) *1983 ASCD 1983 yearbook: Fundamental curriculum decisions*, 53-67. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kellow, J. T. & Willson, V. L (2001). *Consequences of (mis)use of the Texas assessment of academic skills (TAAS) for high-stakes decisions: A comment on Haney and the Texas miracle in education*. ERIC Clearinghouse on Assessment and Evaluation. Available online from: <http://ericae.net/pare/getvn.asp?v=7&n=24>
- Killion, J. (2000). Effective staff learning must be linked to student learning. *Results*, May. Available online from: www.nscd.org/library/results/res5-00kill.ntml
- Klenzschy, M., Garrison, L. & Amaral, O. M. (1999). *Valle Imperial project in*

- science (VIPS) four-year comparison of student achievement data 1995-1999*. Calexio, CA: Educational Research Institute.
- Kluger-Bell, B. (2000). *Recognizing inquiry: Comparing three kinds of hands-on teaching techniques*. In R. Brown (Ed.), *Foundations: A monograph for professionals in science, mathematics, and technology education. Inquiry: thoughts, views, and strategies for the K-5 classroom* (pp. 39-50). Arlington, VA: National Science Foundation.
- Krejcie, R. V. and Morgan, D. W. (1970). In Gay, L. R., (1996). *Educational research: competencies for analysis and application, (5th Ed.)* (p. 125). Columbus, OH: Merrill.
- Landis, C. (1996). *Teaching science in the field*. Columbus, OH: ERIC Clearing House for Science, Mathematics, and Environmental Education. Available online from: www.ed.gov/databases/ERIC_Digests/ed402154.html
- Loucks-Horsley, S. & Pratt, H. (1993). *Implementing a science curriculum for middle grades: Progress, problems, and prospects*. In G. M. Madrazo, Jr. & L. L. Motz (Eds.), *Sourcebook for science supervisors (4th ed.)*, (pp. 61-72). Washington, D.C.: National Science Teachers Association.
- Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.
- Luft, J., (1999). Challenging myths. *The Science Teacher* 27(1), 40-43.
- Martinello, M. L. (1998). Learning to question for inquiry. *The Educational Forum*, 63, 164-171.

Marzano, R. J., Pickering, D. J. & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Aurora, CO: Mid-Continent Research for Education and Learning Institute. Available online from:

www.ascd.org/readingroom/books/marzano01book.html

Means, B. & Olson, K. (1999). *Technology's role in student-centered classrooms*.

In H.C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (297-318). Berkley, CA; McCutchan Publishing.

Morrissey, M., Cowan, D., Leo, T. & Blair, L. (1999). Renewing teachers, reforming schools through professional learning communities. *SEDLetter*, 11(2), 8-11.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington, D. C.: Department of Education.

National Research Council (1997). *Introducing the national science education standards*. Washington, D. C.: National Academy Press.

National Research Council (1996). *National science education standards*. Washington, D.C.: National Academy Press.

National Research Council (1999a). *Designing mathematics or science curriculum programs*. Washington, D. C.: National Academy Press.

National Research Council (1999b). *Selecting instructional materials*. Washington, D. C.: National Academy Press.

National Research Council (2000). *Inquiry and the national science education*

standards: A guide for teaching and learning. Washington, D. C.: National Academy Press.

National Science Teachers Association (1990) *An NSTA position statement: Laboratory science.* National Science Teachers Association: Available online from: www.nsta.org/handbook/labsci.asp

Oakes, J., Wells, A. S., Yonezawa, S. & Ray, K. (1997). *Equity lessons from detracking schools.* In A. Hargraves (Ed.), *1997 ASCD yearbook: Rethinking educational change with heart and mind*, pp.43-72. Alexandria, VA: Association for Supervision and Curriculum Development.

O'Sullivan, C. Y. & Weiss, A. R. (1999). Student work and teacher practices in science: A report on what students know and can do. *The nation's report card: National assessment of educational progress.* National Center for Education Statistics.

Pedersen, S. (February, 2002). *Effect sizes and "what if" analyses as alternatives to statistical significance tests.* Paper presented at the meeting of the Southwest Educational Research Association, Austin, TX.

Peng, C. J., So, T.H., Stage, F. K. & St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999. *Research in Higher Education.* 42(4) 259-271. Association for Institutional Research.

Popham, W. J. (2001). Confounded causality. *The truth about testing: An educator's call to action.* Association for Supervision and Curriculum Development. Available online from:

www.ascd.org/readingroom.books.popham2001_chapter4.html

Rankin, L. (2000). Lessons learned: Addressing common misconceptions about inquiry, *Foundations: Inquiry thoughts, views and strategies for the k-5 classroom*. (33-38). Arlington, VA: National Science Foundation.

Ravitch, D. (2000). *Left back*. New York: Simon & Schuster.

Schmidt, P. R. (1999). KWLQ: Inquiry and literacy learning in science. *Reading Teacher*, 55(7), pp. 89-92. Available online from:
<http://ericae.net/ericdc/EJ581466.htm>

Schrag, P. (1997). The near-myth of our failing schools. *Atlantic Monthly*, October, pp. 72-97.

Sevenair, J. P. & Burkett, A. R. (2001). *Materials to accompany a presentation on multiple-choice questions, examinations, and test banks*. Xavier University of LA. Available online from:
<http://webusers.xula.edu/jsevenai/objective/objective.html>

Sparks, D. & Hirsh, S. (1997). *A new vision for staff development*. Oxford, OH: National Staff Development Council.

St. John, M. (1999). *The principals of educational reform: Supporting mathematics and science teaching in your school, a handbook for elementary and middle school principals*. Inverness Research Associates. Available online from: www.inverness-research.org/reports/

St. John, M. (2000). *The Exploratorium's Institute for Inquiry four cornerstone claims: A summary of evaluations and findings 1996-2000*. Inverness Research Associates. Available online from: www.inverness-

research.org/reports/

Teddlie, C. & Reynolds, D. (with Creemers, B., Scheerens, J., & Townsend, T.) (1999). *An introduction to school effectiveness research*. In Teddlie, C., & Reynolds, D., (Eds.), *International handbook of school effectiveness research: An international survey of research on school effectiveness* (3-7). London: Falmer Press.

Texas Education Agency, (1998). *Texas essential knowledge and skills*.

Available online from: www.tenet.edu/teks/science

Texas Education Agency, (1999). *TAAS technical digest*. Available online from:

www.tea.state.tx.us/student.assessment/resources/techdig/index.html

Texas Education Agency, (2001). *An overview of the academic excellence indicator system for the state of Texas*. Available online from:

www.tea.state.tx.us/perfreport/aeis/about.aeis.html

Texas Education Agency (2002a). *TAKS information booklet: Science, grade 10*.

Available online from:

www.tea.state.tx.us/student.assessment/taks/booklets/index.html

Texas Education Agency (2002b). *Spring 2001 performance standards – English version*. Available online from:

www.tea.state.tx.us/student.assessment/scoring/pstandards/perfst01.html

Texas Education Agency (2002 c). *2001 district AEIS report*. Available online from: <http://www.tea.state.tx.us/perfreport/aeis/>

Tobin, K. (1999). *Social-constructivist perspectives on the teaching of science*. In H.C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice*

- and research* (269-284). Berkley, CA; McCutchan Publishing.
- Vornberg, James A. (Ed.) (1998). *Texas public school organization and administration: 1998*. Dubuque, IA: Kendall/Hunt Publishing
- Vygotsky, L. (1978) *Mind in society*. In National Research Council (1999c). *How people learn*. Washington, D. C.: National Academy Press.
- Wang, J. (1996). Reassessing a learning gap: A comparative study of student science achievement in the United States and China. *Phi Delta Kappan*, Nov, 234-239.
- Wertz, E. (1999). *National education goals: Lessons learned, challenges ahead*. National Education Goals Panel. Available online from: www.ed.gov/pubs
- Williams, B., (1999). Diversity and education for the 21st century. In D. D. Marsh (Ed.), *ASCD yearbook 1999: Preparing our schools for the 21st century* (89-114). Alexandria, VA: Association for Supervision and Curriculum Development.